

THIS WEEK

EDITORIALS

REMAINS Court asked to determine home of Italian skull **p.462**

WORLD VIEW Governments must learn to love controlled trials **p.463**

HOT LEGS Temperature of skink eggs lengthens limbs **p.465**



Counting the cost

As more and more of its ocean-sciences budget is eaten up by operational and maintenance costs, the US National Science Foundation should learn to take a long view when investing in major projects.

One should never, as the saying goes, look a gift horse in the mouth. So when the US National Science Foundation (NSF) was handed stimulus cash after the collapse of the wider economy, it is not hard to see why agency bosses rushed ahead with funding some shiny new projects. But now the hungry horse is their responsibility and it is gobbling from the agency's shrinking nosebag. Something has to give.

A decade ago, things were looking up for the science-funding agency. Budgets had been rising steadily, thanks to a supportive Congress. In 2007, President George W. Bush signed the America COMPETES Act into law, which singles out the NSF for special investment in innovation research. And even after the US economy nosedived in 2008, there was a silver lining for the NSF: an extra \$3 billion in stimulus funding from the government's economic-recovery package.

By law, the NSF was obliged to spend the stimulus windfall quickly, so it naturally looked to inject cash into projects that were 'shovel-ready' — those that had already been designed and were just waiting for investment to get started. Of the many things that the agency did with the cash, it chose two large projects in its ocean-sciences division. A group hoping to build an Alaskan research vessel received more than \$100 million to begin construction, as did a project looking to establish a network of ocean observatories in the waters surrounding the Americas (see page 480).

Even then, the NSF should have anticipated that the big budgets would not last and planned accordingly. It did not, and now faces the reality of the aftermath of all that spending. Once the pot of money allocated to construction has gone, the agency must start to pay operational costs for these expensive projects. Both the Alaskan vessel and the Ocean Observatories Initiative (OOI) are set to come online in 2015, and the ocean-sciences division will foot the bill. In a presentation to the National Science Board last month, division director David Conover warned that the division is already spending more than half its money on maintaining facilities — at the expense of core science projects. And that percentage of facilities costs is only expected to grow.

That could hurt another long-standing part of the ocean-sciences division — scientific ocean drilling, in the shape of the drilling ship *JOIDES Resolution*. Faced with growing facilities demands, the NSF is considering cutting the amount it spends on the *Resolution* each year, such that its time at sea might shrink from the eight months a year of science it does at present — which is, in turn, less than the 12 months a year it worked a decade ago (see page 469).

Ocean drilling has already absorbed cut after cut; it must be spared complete dismantling. The *Resolution* is a hugely successful science programme, one that continues to yield multiple papers in top academic journals each year, more than four decades after scientific ocean drilling began. It is also highly international; in the past decade, 758 scientists from 23 countries have sailed aboard the *Resolution* under the mantle of the Integrated Ocean Drilling Program. One-quarter of those

were graduate students, of whom nearly half were women.

The NSF faces difficult choices, as do other cash-strapped funders around the world. In the case of the NSF's ocean sciences, it should choose to pay to keep the *Resolution* working. It has little leeway on the expensive and untested OOI, which has been mandated by Congress. (Although, notably, other countries, such as Australia and Canada,

"When money gets tight, some dreams simply have to be delayed."

have managed scientifically useful ocean observatories on a smaller, more affordable scale.) That leaves a decision to be made on the country's ageing academic research fleet.

Even in these tight budgetary times, the NSF is about to embark on another major construction push in the ocean sciences: it is looking to build as many as three regional research vessels. These are much-needed replacements that would study algal blooms, ocean acidification, fisheries impacts and other science of great societal relevance. But they are coming at just the wrong time and should be postponed.

As it awaits confirmation of a new director, the NSF would do well to reconsider the way it builds long-term strategy. Building big, shiny facilities is all well and good in times of plenty. When money gets tight, some dreams simply have to be delayed. With politics, as with horses, there is no sure thing. ■

Time for change

Angela Merkel needs to tackle the issue of Germany's uneven university funding.

With her triumph in the German parliamentary elections on 22 September, Angela Merkel's popularity has reached new heights. Her bloc — the Christian Democratic Union and its Bavarian sister party — took 41.5% of the vote, just five seats short of an absolute majority and almost 8% more than her share in the 2009 election. But as the Free Democratic Party, her junior coalition partner in the last government, failed to win the required 5% of votes and will no longer be represented in parliament, Merkel must seek a new political partner. A grand coalition with the Social Democrats, who won 25.7% of votes, seems the most likely option. It could be a good one for science as well.

Merkel no doubt owes her victory to Germany's economic stability and her firm stance on the euro crisis, which has made her the pre-eminent political figure in Europe. Her government has also cut German unemployment by almost 40% since 2005, to 6.8%. And Merkel has benefited from her decision to pull the plug on nuclear

energy by 2022 in the wake of the 2011 accident at the Fukushima Dai-ichi nuclear-power plant in Japan. The cost and technical challenges of the *Energiewende*, the move to a non-nuclear, low-carbon energy system (see *Nature* 496, 156–158; 2013), will dominate her third term in office. As will coping with the welfare and health pressures brought about by an ageing population.

A lot of good science will be needed to meet these challenges. Wisely, the government has increased research and technology expenditure by some 60% since 2005 (see *Nature* 501, 289–290; 2013). Today, Germany's science landscape is more diverse, more competitive, better funded and less parochial than at any time since the Second World War. Many Max Planck Institutes offer terms and conditions that few other places in the world can match. National research centres, such as the Alfred Wegener Institute for Polar and Marine Research, are among the leading hubs in their fields, and the model of the Fraunhofer Society, which promotes applied research in conjunction with industry, is now being copied by the United Kingdom. All these organizations, as well as the DFG — Germany's central grant-giving agency for university research — have benefited from the Pact for Research and Innovation, which has given them generous budget increases over the past few years. Merkel has promised to continue this pact beyond 2015, which would guarantee them budget increases of 5% each year.

But not all is rosy. German scientists are at a disadvantage in stem-cell research compared with countries such as Sweden or the United Kingdom. German law prevents the importation or use of any human embryonic stem cells except those created for research before 1 May 2007. The Free Democrats are the only party to have backed more liberal stem-cell rules in the past, and their absence from parliament makes a revision of the law unlikely.

Life could also be better for some plant biologists. Research on genetically modified (GM) crops has all but stopped owing to public hostility and a lack of political support. Since 2005, all experimental releases of GM plants have had to be registered to give their exact location and time of planting. This has allowed opponents to destroy nearly every field trial. As a result, for the first time in 20 years, there were no GM field trials in Germany this year.

“The priority for Merkel should be to strengthen the country's underfunded universities.”

The government must rethink its anti-GM policies, which are not supported by any scientifically credible risk assessment. With scientific literacy in the basics of plant breeding and genetics at a low level in Germany, public debate about the field is wide open to quacks and ideologists.

But the first priority for Merkel, as *Nature* has called for previously, should be to strengthen the country's relatively underfunded universities. The universities are the responsibility of the country's 16 states — a funding model that has proved incapable of supporting powerhouses to rival the likes of Harvard or Oxford. The €4.6-billion (US\$6.2-billion) Excellence Initiative, jointly funded by central government and the states, has injected some much-needed federal money into the university system. It would take just a two-word constitutional change to allow the government permanently to support state-funded universities — or even to create national research universities similar to Switzerland's Federal Institutes of Technology. In the past, the second chamber of parliament has blocked such an amendment, but it will find it harder to keep up its resistance if Germany ends up being ruled by a grand coalition. ■

Homes for bones

A dispute over the skull of an Italian cheese thief highlights the enduring debate over repatriation.

It is understandable that indigenous communities want to take control of their cultural history. In the past few decades, Native Americans, Australian aborigines, Australian Torres Strait islanders and other groups previously colonized and suppressed by European nations have engaged museums in a rightful debate over whether ancestral bones should be returned to their communities of origin.

The Smithsonian Museum in Washington DC began to return some Native American bones in the late 1980s. And in April this year, the German Museums Association formally agreed that human remains collected as part of a violent conflict should be repatriated. Museums are cautious, however. They recognize the dangers of breaking up scientifically important collections — which have over the decades and centuries become part of world heritage in their own right — if claims to ownership are not clear-cut.

A bizarre case on this sensitive theme is building to a conclusion in Italy. Almost a year ago, a judge in the southern region of Calabria ruled that the skull of a man called Giuseppe Vilella should be returned (“for decent burial”) to the small Calabrian town of Motta Santa Lucia, where Vilella was born around 1801. The skull is a key exhibit in the Cesare Lombroso Museum of Criminal Anthropology in Turin, northern Italy. The University of Turin, which owns the museum, has appealed the ruling and a decision is expected in December.

The case is a one-off, but it highlights a pressing need for greater legal protection of Italy's wealth of historically important scientific objects. In 2004, a law extended protection of the country's remarkable artistic and archaeological heritage to scientific collections in public museums. But Motta Santa Lucia's claim would take the skull out of

the collection — and into legal limbo.

Little is known about Vilella other than that he ended his days in a prison near Pavia in northern Italy, where he had been held for stealing goats and cheese. After he died in 1864, Lombroso, then a professor of forensic medicine at the University of Pavia, acquired his skull and noted an abnormal hollow on the inside back surface. This set Lombroso on course to develop a notorious theory that criminality was an inborn characteristic recognizable through particular anatomical features. He went on to collect hundreds of other skulls to back up this theory. It proved incorrect, but does demonstrate Lombroso's revolutionary willingness to consider that behaviour could be influenced by brain biology.

The judge's ruling is frustrating. Without calling on scientific expertise — a tendency of Italian judges that has been increasingly criticized (see *Nature* 491, 7; 2012) — he said that because Lombroso's theory was known to be wrong, there could be no justification for keeping the skull in a museum.

The inhabitants of Calabria can hardly be considered a suppressed indigenous population. But a tiny political group called the Neo-Bourbon Movement (*Movimento Neoborbonico*) thinks that the analogy holds. Whereas conventional history considers the creation of the Kingdom of Italy in 1861 to have been a liberation of the south by the north, the Neo-Bourbon Movement views it as an invasion that harmed the southern cultural identity. The movement persuaded the mayor of Motta Santa Lucia to bring charges against the Lombroso museum.

The 2004 Italian cultural-heritage law is set to be updated soon, providing a perfect opportunity to extend protection explicitly to individual scientific objects. This would close a legal loophole and sensitize judges to the true value of the objects, which, like artworks, should not in most circumstances be destroyed or lost to the public.

In the meantime, the Lombroso museum is allowed to keep Vilella's skull on display. The bones await their fate on a shelf just a few metres away from a cabinet that holds the entire — less sensitive — skeleton of Lombroso himself. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunjv



Behavioural insights are vital to policy-making

Governments should embrace the scientific approach and use controlled trials to test the impact of policies on people's behaviour, says Olivier Oullier.

Policy-making by governments affects the behaviour of large numbers of people, sometimes millions. So why is such a key task often left to economists and lawyers, who may have little in-depth understanding of how people really behave? And why are the behavioural psychologists and neuroscientists who have valuable expertise usually consulted last, if at all?

Some politicians recognize this problem, and have tried to address it in the past few years. Leading the way is the British government under Prime Minister David Cameron, who established a Behavioural Insights Team (BIT) within the influential Cabinet Office shortly after he was elected. This summer, the White House Office of Science and Technology Policy began hiring people for the US equivalent. And on 30 September, some 300 people — including leading executives in governments, businesses, non-governmental organizations and academia — will gather in Brussels to discuss how behavioural insights can inform policy-making.

The conference has been organized by the European Commission's Directorate General for Health and Consumers, which relies on the insights of behavioural scientists it has recruited in recent years. For instance, the choice of the next library of pictorial and text warnings on cigarette packets will be based on tests of their effectiveness, such as tracking people's eye movements when they view them. The European Commission introduced a set of such warnings in 2005, but they were tested only with highly biased declarative methods, such as surveys and self-reporting.

The lesson here is that if a governing body such as the European Commission — infamous for its bureaucracy and resistance to change — can see the benefits of behavioural insights and alter its way of informing policy, then every government should follow its lead.

At the heart of this approach are the randomized controlled trials that are already common in medical research. Similar trials of public policy are crucial because they use a control group — a fraction of the population to which the new policy is not applied. This might sound strange, but monitoring such a non-intervention group is the only way to know whether a change in behaviour is down to the policy being trialled.

The British government is streets ahead on this. When the BIT was founded in 2010, the French Prime Minister's Centre for Strategic Analysis was already running a programme to chart the benefits of using behavioural and brain sciences to inform public policy, which I had led for a year. But over the two years that followed, while we published reports with the hope of convincing our administration to give our field studies the green light, the British team was running trials and getting results.

One successful example of a cost-effective use of behavioural insights in policy is a UK study on tax collection. In a 2011 randomized controlled trial of more than 100,000 people, some people received payment-request letters that had been tweaked to say that most UK citizens pay their taxes on time. Compared with control letters, the trialled policy produced a 15% increase in repayment rate. The British government estimates that a national roll-out of the policy would provide around £30 million (US\$48 million) of extra revenue each year. Not bad for a smart use of social psychology.

Randomized controlled trials can help governments to choose the best strategy, to spot errors and thus to save a lot of money. The private sector has long understood that. Take OPower, a company headquartered in Arlington, Virginia, that promotes the sustainable consumption of energy. It reports that its clients have saved more than 2.8 billion kilowatt hours of energy as a result of its clever blend of trials, insights from social psychology and behavioural economics — also known as nudges, after the title of a 2008 best-selling book. The company found that the best way to get people to save energy was not to make them feel guilty about the environment or to promise cheaper bills, but to provide households with an easy way to compare their energy consumption with that of their neighbours. Social comparison and peer pressure work better than any lecture on how one should behave.

Many of us who work in the field of behavioural change have received educational training and research grants paid for by public funds.

Yet the private sector seems to be much more interested in our skills than are public institutions, despite the high return that we could provide on investments in behavioural-sciences research. During these difficult economic and social times, public organizations can no longer afford — financially and socially — not to test policies before they are applied.

I am sure that most readers, regardless of where they live, could share examples of failed policies that have seen public money wasted on education, health, law, transport and taxes — all because no one cared to assess how citizens might behave when new measures are implemented. No one would accept that a new drug would be developed only by economists and lawyers and launched without the proper trials. We should not tolerate this in policy-making either. ■

Olivier Oullier is a professor of behavioural and brain sciences at Aix-Marseille University, Marseilles, France. He is former head of the neuroscience and public-policy programme at the Centre for Strategic Analysis of the French Prime Minister, and a World Economic Forum Young Global Leader.
e-mail: olivier@oullier.fr

**PUBLIC
ORGANIZATIONS CAN
NO LONGER
AFFORD
NOT TO TEST POLICIES
BEFORE THEY ARE
APPLIED.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/8gyo9e

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

PSYCHOLOGY

Young people resist bad news

Children and teenagers get better at learning from bad news as they grow older, according to a study by Christina Moutsiana and her colleagues at University College London.

Fifty-nine people aged 9 to 26 were asked how likely they thought they were to experience something bad, such as a car accident or burglary. The team then showed the volunteers actual probabilities for such events and asked them again.

The younger the volunteer, the worse he or she was at learning that a misfortune was more likely than previously believed. But if the chance of a bad experience was less likely than initially predicted, the good news got across to all ages. These results could explain why teens are inclined towards risky behaviour in spite of warnings. *Proc. Natl Acad. Sci. USA* <http://doi.org/nv5> (2013)

GENE REGULATION

RNA boosts genes on a loop

Bits of a gene that don't code for protein can regulate that gene's expression as a loop of RNA.

Ling-Ling Chen, Li Yang and their colleagues at the Shanghai Institutes for Biological Sciences in China examined a class of 'introns' — sequences that reside in the protein-coding portion of genes but get snipped out of the RNA template before the protein is made.

Some introns formed circles that were abundant in the nucleus and, unlike other reported circular RNAs, did not associate with small regulatory RNAs. Reducing

levels of the circular intronic RNAs, many of which were unique to human cells, reduced expression of the parent genes. *Mol. Cell* <http://doi.org/nv6> (2013)

NEUROSCIENCE

Lit neurons show sleep patterns

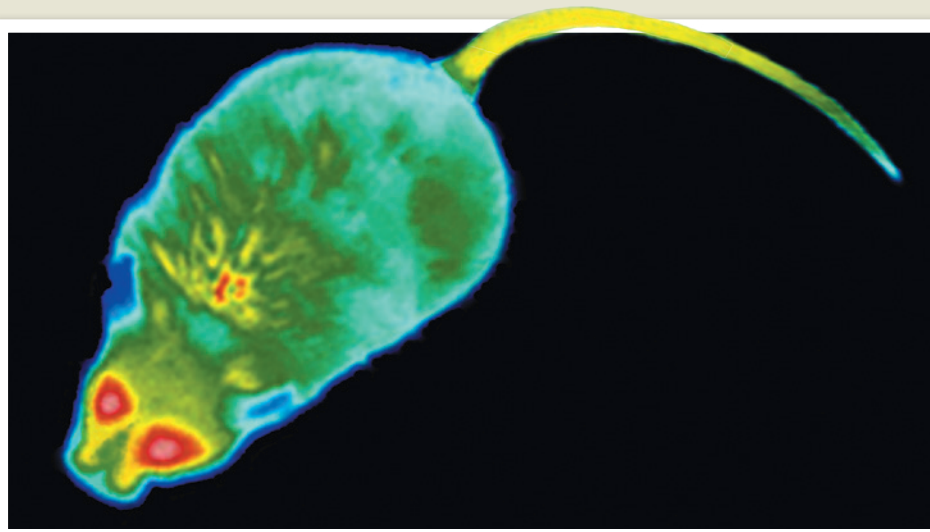
Experiments within living brains have revealed key neurons that control a stage of deep slumber called rapid-eye movement (REM).

In mouse studies, scientists led by Antoine Adamantidis of McGill University in Montreal, Canada, engineered a group of neurons so that they could be

controlled by external pulses of light in the brain region linked to REM sleep. The neurons make a peptide called melanin-concentrating hormone (MCH).

Activating these cells with light increased the likelihood that mice entered REM sleep and extended its duration by 47%. Silencing the neurons reduced the quality of REM sleep (as measured by brain waves called theta oscillations) but did not shorten it.

The neurons may increase the stability of REM sleep by suppressing neighbouring neurons linked to arousal and wakefulness, the authors say. *Nature Neurosci.* <http://doi.org/nw7> (2013)



AMY WARNER

METABOLISM

Thyroid disease tips body's thermostat

Some unexpected effects of thyroid hormone on blood vessels could explain why people with thyroid disorders often feel too hot or too cold.

Jens Mittag of the Karolinska Institute in Stockholm and his colleagues found that mice with a mutation in a thyroid-hormone receptor burn fat at a higher rate than normal, but show no increase in their body temperatures.

Thermal imaging revealed that the mutant

mice lost more heat than usual through their tails because the tail arteries were not constricting properly. This, in turn, boosted the rate at which the mice burned brown fat — an effect that was reversed when the mice were given a drug that stimulates blood-vessel constriction.

Proc. Natl Acad. Sci. USA <http://doi.org/nv7> (2013)

MICROBIOLOGY

Termite turns to its dung for defence

A species of termite protects itself from fungal infection by using the faeces lining its nest to cultivate bacteria.

Colonies of *Coptotermes formosanus*, a subterranean termite, are constantly exposed to pathogens such as the fungus *Metarhizium anisopliae*. For years, researchers have been trying to exploit the fungus to stop *Coptotermes* from chomping through wooden structures, but the insects have evolved several ways of resisting the pathogen.



Thomas Chouvenec and his colleagues at the University of Florida in Fort Lauderdale report another such defence mechanism: the mixture of wood pulp and faeces that lines *Coptotermes*' nests (pictured) supports bacteria that have antimicrobial activity, including a *Streptomyces* species that is effective against *Metarhizium*. The bacterium gave termites a considerable survival advantage in soil that harboured the fungus.

Interfering with this mutually beneficial insect–bacterium relationship might aid pest control, and termite faeces could also be an untapped source of medicines that protect against infection, the authors say. *Proc. R. Soc. B* 280, 20131885 (2013)

EVOLUTION

Cool eggs make for longer legs

Incubating eggs of the tropical skink (*Carlia longipes*; pictured) at different temperatures results in hatchlings with legs adapted for different terrains.

Brett Goodman, now at the University of Adelaide in Australia, and his colleagues caught female skinks living in a mixed habitat of forest and

rocks. These lizards lay clutches of two eggs, and the researchers split 76 sets of sibling eggs to incubate at temperatures reflecting either the forest floor (about 28.5 °C) or rock surfaces (about 23.5 °C).

Eggs incubated at the cooler temperature hatched into long-legged skinks that could sprint and climb more quickly than their warm-incubated siblings. When given a choice, the cool-incubated skinks spent more time in rocky areas, whereas skinks incubated at the warmer temperature had shorter legs and preferred leaf litter.

Am. Nat. <http://dx.doi.org/10.1086/673299> (2013)

PHYSICAL CHEMISTRY

Vitamin coating blocks static

The familiar phenomenon of static electricity can cause serious problems, such as explosions and equipment damage.

A team led by Bartosz Grzybowski at Northwestern University in Evanston, Illinois, has found an explanation for how static electricity builds up, along with a simple way to dissipate it.

The researchers examined magnetic and electric patterns on polymer surfaces using specialized forms of atomic force microscopy. This revealed that clumps of charged particles owe their stability to chemical species known as radicals — reactive molecules with unpaired electrons.

Further experiments showed that these radicals could be mopped up by surface coatings, including vitamin E and other antioxidants. Coated polystyrene beads quickly lost their static charge, and a coated transistor resisted damage



COMMUNITY CHOICE

The most viewed papers in science

ANIMAL MODELS

Lab rats reveal genetics of selection

HIGHLY READ
on www.cell.com
in September

DNA sequencing is illuminating the genetic underpinnings of 27 rat strains bred to study human diseases.

Since the early 1900s, researchers have selected and bred rats displaying traits such as hypertension and behavioural disorders, producing hundreds of distinct strains. Researchers led by Timothy Aitman at Imperial College London sequenced the complete genomes of several common strains of laboratory rat, revealing more than 13 million genetic variants as well as evidence for the evolutionary pressures behind them. A strain known for its cardiovascular disorders, for example, showed specific selection for gene clusters that are involved in regulating blood pressure and oxidative stress.

Further analysis of how artificial selection created rat models of disease should point to genetic variations and mechanisms that contribute to human maladies.

Cell 154, 691–703 (2013)

DIAGNOSTICS

Virus revealed by host response

A straightforward test of human immune responses can show whether flu-like symptoms arise from bacterial or viral infection.

A team led by Christopher Woods and Geoffrey Ginsburg of Duke University in Durham, North Carolina, developed an assay based on a technique called RT-PCR, which can measure gene expression in blood samples.

The assay monitors roughly 30 human genes that have differing activities in bacterial and viral infections. When used on 102 people who came to a hospital with fevers and respiratory symptoms, the test identified a viral infection with 94% accuracy and could exclude a viral cause with 89% accuracy.

Sci. Transl. Med. 5, 203ra126 (2013)

NATURE.COM

For the latest research published by Nature visit:

www.nature.com/latestresearch

CLIMATE

Waiting to reduce emissions is costly

If international efforts to reduce carbon emissions are delayed by another 15 years, the initial costs of limiting global warming could triple.

Gunnar Luderer at the Potsdam Institute for Climate Impact Research in Germany and his colleagues used a model to study 285 mitigation scenarios. They calculated the costs incurred within the first few years after implementing policies to hold temperatures to 2 °C above pre-industrial levels.

If such policies come into force in 2030 or later, energy prices are likely to shoot up by 80% and global economic growth might decrease by 7% in the decade following implementation.

Broad emissions reductions need to start in 2015 to keep costs moderate, the authors say. *Environ. Res. Lett.* 8, 034033 (2013)

from an ion gun.

Science 341, 1368–1371 (2013)

For a longer news story on this work, see go.nature.com/ejdp1

SEVEN DAYS

The news in brief

POLICY

Election results

Angela Merkel triumphed in Germany's parliamentary elections on 22 September to secure a third term as the country's chancellor. Federal science spending has increased by 60% since Merkel first came to power in 2005, and she has fought to keep government science expenditure high despite increases in public debt. The Free Democrat Party, which had previously expressed support for easing Germany's tough restrictions on stem-cell research, lost its representation in parliament. See page 461 and *Nature* 501, 289–290 (2013) for more.

Climate closures

Australia's new government is shutting the Climate Commission, an independent agency set up in 2011 to provide information on climate change. But its chair Tim Flannery says the organization will continue, as the Climate Council, with private funding. The government also aims to close the Climate Change Authority, which advises on carbon pricing and emissions cuts, and the Clean Energy Finance Corporation, a 'green bank' that was due to invest Aus\$10 billion (US\$9.4 billion) over the next five years in renewable-energy projects.

US power plants

Regulations proposed on 20 September by the US Environmental Protection Agency would limit carbon dioxide emissions for future fossil-fuel power plants in the United States. To meet the emissions limits, new coal-fired plants would need to install equipment to capture and sequester a portion of their carbon dioxide emissions — a requirement

that utility companies argue would effectively halt future coal-plant development. The agency is preparing separate regulations, due in June 2014, to govern existing power plants. See go.nature.com/c2m4tn for more.

Russian vote

On 18 September, Russia's lower house of parliament approved controversial reforms to the Russian Academy of Sciences. The upper house and President Vladimir Putin are widely expected to approve the law. This would see the 436 institutes and 45,000 staff of Russia's primary basic-research organization managed by a new federal agency overseen

by the government and reporting directly to Putin. Critics say that the move will deprive the 289-year-old body of its independence, and halt attempts to revitalize Russia's science system. See go.nature.com/8zqqme for more.

EVENTS

Arctic ice low

The extent of Arctic sea ice reached its minimum for the year on 13 September, when ice cover fell to 5.1 million square kilometres, the US National Snow and Ice Data Center in Boulder, Colorado, said on 20 September. This year's minimum falls short of 2012's record-breaking thaw but is the

sixth lowest since 1979, when satellite monitoring began. See go.nature.com/jnkee3 for more.

RESEARCH

Census shake-up

The census that has sampled the UK population once a decade since 1801 could be scrapped, according to a consultation launched on 23 September by the Office for National Statistics. The office is considering using information from national databases and smaller annual surveys for the next census, due in 2021. Some researchers are strongly opposed to this option, which they argue



Road test for experimental reactor

A 352-wheel, 800-tonne trailer last week carried a test load (pictured) along the 104-kilometre route through France by which components of the multibillion-euro ITER thermonuclear reactor will travel for assembly in St-Paul-lès-Durance. Seven international partners will contribute parts for the machine, which is designed to show the feasibility of nuclear fusion

as a power source (see *Nature* <http://doi.org/nwq>; 2012). The trailer — 10.4 metres high, 9 metres wide and 33 metres long — mimics the size of the largest components. The convoy travelled at night along roads and bridges that have been modified at a cost of €112 million (US\$151 million). The project is expected to need a total of 230 convoys over the next five years.

ITER ORGANIZATION

would mean losing important details. See go.nature.com/mwepcq for more.

Space mission dead

After a 7.6-billion-kilometre journey, NASA's comet-hunting Deep Impact spacecraft is no more. Project controllers had been unable to communicate with the craft since 8 August, and declared the mission over on 20 September. A computer calendar glitch probably caused the craft to fail (see go.nature.com/yupxju).

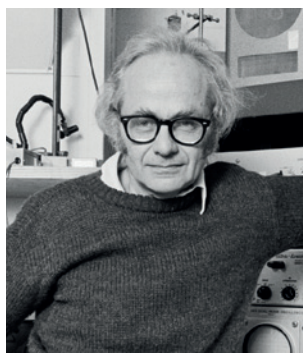
PEOPLE

Smithsonian head

Civil engineer Wayne Clough will retire next year as leader of the Smithsonian Institution in Washington DC, the centre announced on 18 September. Since taking over as secretary in 2008, Clough has raised US\$893 million in private donations for the \$1-billion institute, which hosts some 1,600 fellows and research associates across 19 museums and 9 research centres.

Nobel laureate dies

David Hubel, a Nobel-prizewinning neuroscientist who mapped the brain's visual system, died on 22 September, aged 87. In research covering 25 years, Hubel (**pictured**), based at Harvard Medical School in



Boston, Massachusetts, worked with Torsten Wiesel to show how light is transmitted from retinal cells in the eye to the brain's visual cortex, where it is processed. The duo shared the 1981 Nobel Prize in Physiology or Medicine with Roger Sperry.

GM-rice fallout

Guangwen Tang, a nutrition scientist at Tufts University in Medford, Massachusetts, has been banned by the university from conducting human research for two years owing to ethics violations in a study published last year. Tang tested the health benefits of Golden Rice — genetically modified to produce a vitamin A precursor — on children in China's Hunan province (see *Nature* <http://doi.org/nv9>; 2012). On 17 September, the university upheld the study's results, but reported finding that the trial had violated university and federal regulations, and lacked proper approvals from

Chinese officials. Tufts also said that the participants were not adequately informed about the genetically modified nature of the rice.

Human-rights prize

Omid Kokabee, a physics PhD student jailed in Iran since January 2011, was on 23 September awarded the American Physical Society's Andrei Sakharov Prize, for courage in refusing to work on projects that he thought were harmful to humanity. Kokabee has said that he was pressured to cooperate in an Iranian military nuclear programme. This is the first time that the human-rights prize has been awarded to a person in prison. See go.nature.com/ic77ff for more.

FUNDING

US research boost

The US government has allocated US\$53 million to create 14 Tobacco Centers of Regulatory Science in the next year, the Food and Drug Administration and the National Institutes of Health (NIH) announced on 19 September. More than \$273 million may be invested over the next five years to help the programme to develop and evaluate regulations for tobacco products. The NIH last week also announced some \$45 million in awards

COMING UP

27 SEPTEMBER

In Stockholm, the Intergovernmental Panel on Climate Change releases a summary of its fifth assessment of the basic scientific evidence for climate change. The panel publishes its full report on 30 September.

to study early interventions for Alzheimer's disease. Among the selected projects is a \$33-million trial to test preventive treatments in at-risk adults (see *Nature* **489**, 13–14; 2012).

Cancer cash

On 21 September, philanthropists Phil and Penny Knight announced that they would donate US\$500 million to cancer research at Oregon Health & Science University in Portland — but only if the university can raise another \$500 million from other sources within two years. The money would support science at the university's Knight Cancer Institute, to which the couple gave a \$100-million donation in 2008. Phil Knight is co-founder of the sportswear company Nike.

FACILITIES

Genome centre

On 19 September, the New York Genome Center officially opened the doors of its first dedicated research facility, located in Manhattan. The consortium of 12 research and health-care institutions was created three years ago as part of New York city's push to become a biomedical-research hub. Its aim is to provide non-profit genome services and to foster collaborations between academia and industry. See go.nature.com/htxqji for more.

► **NATURE.COM**

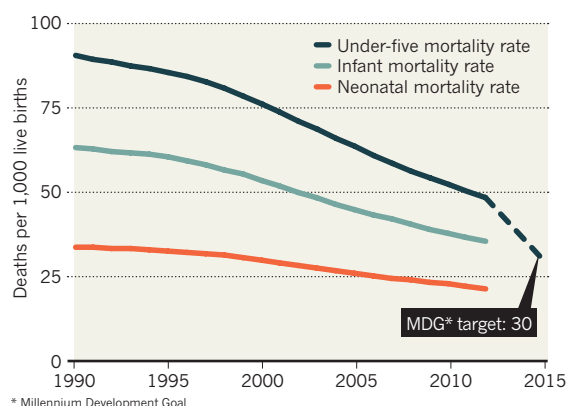
For daily news updates see:
www.nature.com/news

TREND WATCH

In 2012, some 6.6 million children died before reaching their fifth birthday. By contrast, more than 12 million under-fives died in 1990, according to a United Nations report released on 13 September. Although the death rate has been cut by about half, the situation is not improving fast enough to reach a target set for 2015 by the UN Millennium Development Goals initiative (see chart). Worldwide, malnutrition accounts for about 45% of deaths of children under five years old.

CHILD DEATH RATE HALVED

The United Nations says that the world has made encouraging strides in reducing the child mortality rate.



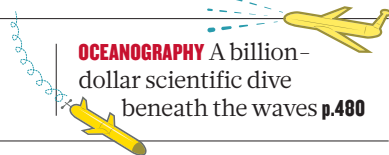
NEWS IN FOCUS

JOURNALS Baffled scientists see findings scooped — by fake authors **p.470**

POLICY Regulatory huffing and puffing over electronic cigarettes **p.473**

PHYSICS Resurrected hope for room-temperature superconductor **p.474**

OCEANOGRAPHY A billion-dollar scientific dive beneath the waves **p.480**



IODP-USIO



US research vessel the *JOIDES Resolution* on an expedition in the Gulf of Alaska.

OCEANS

Drilling hit by budget woes

US funding for research vessel uncertain as international programme reorganizes.

BY ALEXANDRA WITZE

This week's arrival of US research vessel the *JOIDES Resolution* in Busan, South Korea, will mark an important milestone in scientific drilling. On 30 September, the Integrated Ocean Drilling Program (IODP) — the framework that has governed the extraction of geological cores from the sea floor for the past decade — will cease to exist. The next day, it will be reborn for another decade with a new name and blueprint, but the same old problem: how to pay for what is arguably the most successful international research collaboration ever.

The future of the *JOIDES Resolution* is particularly in doubt, which is a source of consternation for the geoscientists who depend on her services. For the past decade, the ship has been a workhorse of scientific ocean drilling, racking up 34 expeditions under the IODP. By contrast, vessels run by other countries have managed just 18 (see 'Drilling down').

But the ocean-sciences division of the US National Science Foundation (NSF), which funds the *JOIDES Resolution* from its headquarters in Arlington, Virginia, has run into a perfect storm of financial commitments. Rising costs for new facilities, coupled with double-digit budgetary cuts in core science programmes over

the past two years, have led the NSF to consider slashing the money available for the vessel.

Rumours of the impending change have caused uproar among US geoscientists. They have sent a flood of letters to the NSF to highlight the valuable research conducted through ocean drilling — the only way to access continuous geological records older than about a million years. International collaborators such as Brazil and China that bring money to NSF coffers through membership fees now face the prospect of less time at sea.

The NSF's head of ocean sciences, David Conover, points out that other programmes in his division are also suffering. "Everybody ▶

► wishes we had more money to give," he says. In November, he will appear before the National Science Board with options on how the *JOIDES Resolution* might operate after its contract with the NSF ends in September 2014. The board will vote to authorize some level of funding, which will dictate how much research the ship can do. It currently costs about US\$65 million a year to conduct roughly four expeditions.

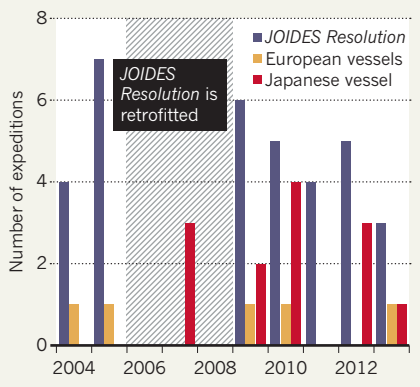
The United States has led the field of scientific ocean drilling since 1968, and today 26 countries share several expedition vessels under the IODP. The programme's other flagship is Japan's *Chikyu*, a \$540-million vessel with 'riser' technology that allows it to drill much deeper than the *JOIDES Resolution* can.

Scientists aboard the *Chikyu* have drilled into the fault that caused the deadly 2011 earthquake near Tohoku, Japan, and have investigated a dangerous fault zone closer to Tokyo, to probe the mechanics of these killer hazards. But the *Chikyu*, like the *JOIDES Resolution*, has faced higher operating costs than expected; it spends about five months a year on science, and the rest of its time on mostly industry work, to help to pay the bills. At an April workshop in Tokyo, *Chikyu* scientists laid out an ambitious wish list of projects, from exploring an ancient Mediterranean dry period to drilling into Earth's crust-mantle boundary, that would easily fill 12 months a year if Japan's government could afford them.

Other parts of the IODP effort are run by the European Consortium for Ocean Research Drilling, which operates 'mission-specific' platforms — hiring ships that are matched to

DRILLING DOWN

Budget pressure has reduced the number of missions undertaken by US drilling ship the *JOIDES Resolution*.



each expedition. The group has scrambled to find money to operate roughly one mission a year, and currently has a ship drilling in the Baltic Sea to explore past glacial cycles.

But these long-standing arrangements are being shuffled. In the rechristened, reorganized 'International Ocean Discovery Program', the United States, Japan and Europe will fund their own platforms directly instead of combining funds in a central pot of about \$200 million. After years of courting, US officials have also brought in partners such as Brazil and China, which each pay \$3 million annually to the US programme. Many had hoped that this influx of international funds might allow the *JOIDES Resolution* to sail more expeditions each year.

But the NSF's budget woes make that look unlikely. Several years ago, the agency spent \$115 million on retrofitting the *JOIDES Resolution*, but the ship has spent only seven to eight months a year on science since then. "If it can only continue at that level, I think the community would view the glass as half full rather than half empty," says Keir Becker, a marine geologist at the University of Miami in Florida, and head of the new group that will coordinate activities among the US, Japan and Europe.

"If the United States falls over, then all these other countries are going to be left hanging," says Mike Coffin, a marine geologist at the University of Tasmania in Hobart, Australia. "Everyone is extremely concerned about what the United States is doing." (Australia has put in a five-year bid to continue participation in the programme.)

Conover brushes aside rumours that the United States is abandoning scientific drilling. "We are not terminating the programme," he says. But he adds that the NSF's ocean sciences division can only stretch so far. Its budget this year is almost \$343 million, or \$9 million less than the year before. Budget sequestration and other cutbacks have reduced its funding for core science by about 12% over two years.

At the same time, the 35-year-old *JOIDES Resolution* has only so much life left in it. "The community is waking up and seeing that we don't have this tool forever," says Bradford Clement, who oversees the ship's science services at Texas A&M University in College Station. "There is a sense of urgency here." ■

SEE EDITORIAL P.461

PUBLISHING

Mystery over obesity 'fraud'

Researcher baffled after his results appear in bogus paper.

BY DECLAN BUTLER

Ghost writing is taking on an altogether different meaning in a mysterious case of alleged scientific fraud. The authors of a paper published in July (*A. Veziraki et al. Biochem. Biophys. Res. Commun.* <http://doi.org/nxb>; 2013), which reported significant findings in obesity research, seem to be phantoms. They are not only unknown at the institution listed on the paper, but no trace of them as researchers can be found.

The paper, published in the Elsevier journal *Biochemical and Biophysical Research*

Communications (BBRC), is not the kind of prank that journals have encountered before, in which hoaxsters have submitted dummy papers to highlight weaknesses in the peer-review process. The paper's reported findings — that overexpression of two novel proteins in fat cells leads to improvements in metabolic processes related to diabetes and obesity in mice — are, in fact, true.

Too true, in the opinion of Bruce Spiegelman, a cell biologist at Harvard Medical School's Dana-Farber Cancer Institute in Boston, Massachusetts. He says that he has presented similar findings at about six research meetings,

and is preparing to submit them to a journal. He suspects that the *BBRC* paper was intended as a spoiler of his own lab's work.

Now withdrawn, the article lists five authors who are all supposedly from the School of Health Sciences at the University of Thessaly in Trikala, Greece, and is entitled 'Identification of meteorin and metnl as two novel pro-differentiative adipokines: Possible roles in controlling adipogenesis and insulin sensitivity'. Adipokines are proteins secreted by fat tissue that play an active part in such processes as sugar and fat metabolism, inflammation and obesity-related metabolic disorders, including insulin resistance and diabetes.

Spiegelman says that he smelt a rat as soon as he saw the paper. Meteorin and metnl have been little studied, and no previous paper has shown a role for them in obesity. It was therefore suspicious, he says, to see a paper published out of the blue reporting that they were novel adipokines and that their overexpression in adipose cells led to improvements in diabetes and obesity in mice — exactly the same findings as the work he had presented.

On 20 July, he e-mailed Ernesto Carafoli, *BBRC*'s editor-in-chief, to air his concerns.

"The authors on this paper have apparently never published a single academic paper before and they list a non-academic e-mail address," he wrote. "Odder still, upon looking for them on Google, PubMed or on the website of the university they list, there is no mention of any of the authors as being at that university."

Carafoli, along with Elsevier, launched an investigation. Elsevier temporarily withdrew the paper from the journal website on 8 August, and, after the University of Thessaly confirmed that none of the researchers listed on the paper had ever worked there, now intends to withdraw it permanently.

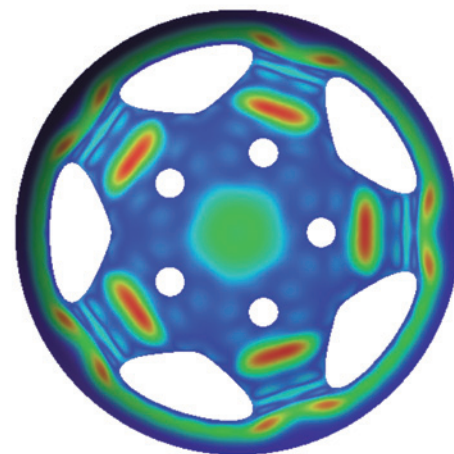
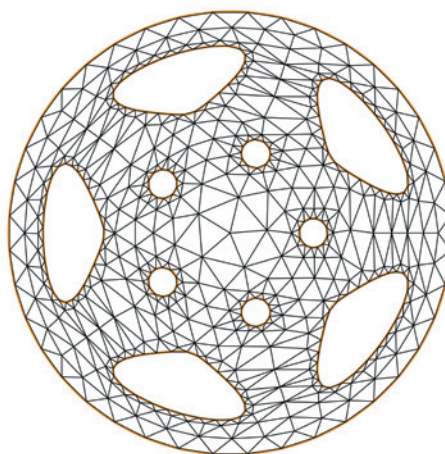
Spiegelman, who works on fat-cell differentiation, is also a co-founder of Ember Therapeutics, a company based in Watertown, Massachusetts, that is developing therapeutics for metabolic disorders. He believes that the paper was intended to hurt him and his lab. Scientific misconduct is usually done for academic gain, but because the authors on the paper seem to be phantoms, they can derive no benefit, he says. He argues that this seems to leave "maliciousness" as the only explanation.

Spiegelman says that he is surprised that the e-mail address of the corresponding author did not prompt the journal to ask for evidence of the authors' institutional affiliations. "The e-mail was a bit strange, and that we could have checked," agrees Carafoli, but nothing else in the paper aroused suspicion. "It was impeccable. The authors were clearly academics," he adds.

The perpetrators also seem to have used Greek surnames similar to those of authentic researchers working in obesity-related research, in what one might speculate was an attempt to fool referees should they search the literature. There are also genuine researchers at the University of Thessaly working in the field of obesity.

Spiegelman, who is certain that the paper is "made up", is keen for there to be a criminal investigation. He says that lawyers have told him that the faked paper represents fraud, not just academic misconduct — a view shared by Carafoli. But Spiegelman says the lawyers also advised that although he might have been the target, there would be little basis for him to sue, whereas Elsevier, BBRC and the University of Thessaly could have grounds to press fraud charges.

Elsevier told *Nature*: "BBRC has been targeted by a scheme to defraud our editors, reviewers and readers with submission of a manuscript with falsified author and institutional information and therefore wholly unverifiable scientific claims. We consider such abuse unethical." It added that it is continuing its investigation and will, with the relevant authorities, "explore the question of whether this constitutes a criminal case of Internet fraud." ■



A surface imaged by a patented method that was licensed to Intellectual Ventures by Caltech.

TECHNOLOGY TRANSFER

Universities struggle to make patents pay

Surfeit of unlicensed intellectual property pushes research institutions into unseemly partnerships.

BY HEIDI LEDFORD

United States patent number 7,023,435 almost didn't happen. The application, which covered a way of imaging a surface, was rejected four times by the US Patent and Trademark Office. But the California Institute of Technology (Caltech) in Pasadena, which filed the patent, fought back — and prevailed in 2005.

Caltech's faith in the hard-won patent was not matched by industry: three years later, no one had licensed the rights to the invention. So in 2008, Caltech exclusively licensed it, along with 50 other patents, to a subsidiary of Intellectual Ventures, a company that has stockpiled 40,000 patents from which it collects US\$3 billion in licensing income. The firm sometimes uses its patents to sue other companies for infringement, yet it rarely develops the inventions described by its intellectual property.

Such patent-assertion entities, sometimes called aggregators, monetizers or 'patent trolls', are questionable homes for university inventions. But in the push to get academic research out of the ivory tower — and to make money — university technology-transfer offices are becoming less choosy about their partners.

"As universities struggle to find revenue sources, one might worry that the monetization industry will be very tempting," says Robin Feldman, director of the Institute for Innovation Law at the University of California Hastings College of the Law in San Francisco.

There are already signs that this is happening, she adds. Last year, she published evidence that 45 universities around the world licensed or sold patents to Intellectual Ventures shell companies (T. Ewing and R. Feldman *Stanford Technol. Law Rev.* 1; 2012).

Intellectual Ventures, headquartered in Bellevue, Washington, chafes at the term patent troll. The company's global head of technology, Patrick Ennis, points to its role in launching three start-up companies, and to deals it has struck with Caltech and other universities to sponsor

"In the end, it just came down to money."

research in exchange for ownership of the resulting patents as evidence of the firm's commercial activities.

Universities often say that the goal of licensing patents is to stimulate the economy by translating publicly funded research into companies and products. But the unstated aim is to make money to fund more research and the technology-transfer office itself, says Melba Kurman, a former technology-transfer officer at Cornell University in Ithaca, New York, who now works as a consultant. The goals are sometimes in conflict. "If the goal is to monetize the patent portfolio, then it would make sense to auction it off to the highest bidder," says Kurman. "But when these patents cover taxpayer-funded research, that is not an acceptable solution."

Finding a bidder at all can be a coup for technology-transfer officers, who are often saddled with patents that are years away ▶

► from practical application. Joy Goswami, assistant director of the technology-transfer office at the University of Delaware in Newark, estimates that only about 5% of patents are licensed at most universities. The rest are a drain on office resources, he adds, because of required maintenance and legal fees.

Earlier this month, at a meeting of the Association of University Technology Managers in Boston, Massachusetts, Goswami led a discussion on how to unload the 95% of patents that remain unlicensed. One option is to use a broker or auction house that specializes in trading intellectual property. It is a controversial solution that some universities are afraid to touch. "When you go to sell a patent, the university loses any ability to ensure that it's being managed in the public's best interest," says Kelly Sexton, head of the technology-transfer office at North Carolina State University in Raleigh, which licenses out a relatively robust 15% of its patents.

Thomas Major, vice-president of IPOfferings, a patent-broker firm based in Boca Raton, Florida, does not understand the hesitation. He spent nine years managing intellectual property at the University of Utah in Salt Lake City, and says that universities would be foolhardy to ignore auction houses. "When I was at the University of Utah, I would have sold those patents in a heartbeat," he says.

Major says that IPOfferings has handled about 20 patents from universities over the past 3 years — roughly 7% of the firm's total business. James Malackowski, chief executive of Ocean Tomo in Chicago, Illinois, a company best known for its patent auctions, says that universities are increasingly coming to him, and represent nearly 20% of his business. Both Major and Malackowski say that their firms can place limits on who can buy or exclusively license a patent. Even so, Major says that at least one of his university clients abandoned that request after seeing how much a patent aggregator was willing to pay. "In the end, it just came down to money."

Such decisions violate the spirit of a 2007 memo endorsed by more than 100 institutions. Offering guidance for ethical patent licensing, it cites the risks of dealing with patent aggregators. Yet the signatories include Caltech and three other universities that have licensed patents to Intellectual Ventures, according to Feldman's research: Duke University in Durham, North Carolina; the University of Florida in Gainesville; and the University of Ottawa in Canada. The universities declined to comment on the patent deals.

Caltech mathematician Peter Schröder, one of the three inventors on US patent 7,023,435, is not too bothered. He would be troubled, he says, if Intellectual Ventures were using his patent to harass other companies, but so far he has not heard of that happening. "It's not giving me heartache," he says. ■

BIOLOGY

Mozilla plan seeks to debug scientific code

Software experiment raises prospect of extra peer review.

BY ERIKA CHECK HAYDEN

When ecologist Carl Boettiger wrote a blog post in June calling for greater stringency in the peer review of scientific software in research papers, he hardly expected to stir up controversy. But in 54 comments on the post, researchers have debated how detailed such reviews should be; one said that it was a "trifle arrogant" of Boettiger, of the University of California at Santa Cruz, to insist that computer code attain his stringent standards before publication.

Now an offshoot of the Internet non-profit organization Mozilla has entered the debate, aiming to discover whether a review process could improve the quality of researcher-built software that is used in myriad fields today, ranging from ecology and biology to social science. In an experiment being run by the Mozilla Science Lab, software engineers have reviewed selected pieces of code from published papers in computational biology. "Scientific code does not have that comprehensive, off-the-shelf nature that we want to be associated with the way science is published and presented, and this is our attempt to poke at that issue," says Mozilla Science Lab director Kaitlin Thaney.

Researchers increasingly rely on computation to perform tasks at every level of science, but most do not receive formal training in coding best practice. That has led to high-profile problems. Some scientists have argued, for example, that the fraudulent findings used as the basis for clinical trials in 2007 would have been exposed much earlier if cancer researcher Anil Potti of Duke University in Durham, North Carolina, had been compelled to publish his data and computer code along with his original papers.

More routinely, incorrect or slipshod code prevents other researchers from replicating work, and can even lead them astray. In 2006, Geoffrey Chang of the Scripps Research Institute in La Jolla, California, had to retract five research papers on crystal structure after finding a simple error in the code he was using, which had been provided by another lab. "That's the kind of thing that should freak any

scientist out," says computational biologist Titus Brown at Michigan State University in East Lansing. "We don't have good processes in place to detect that kind of thing in software."

Mozilla is testing one potential process, deploying the type of code review that is routinely used on commercial software before it is released. Thaney says that the procedure is much like scientific peer review: "The reader looks for everything, from the equivalent of grammar and spelling to the correctness of the logic." In this case, Mozilla opted to examine nine papers from *PLoS Computational Biology* that were selected by the journal's editors in August. The reviewers looked at snippets of code up to 200 lines long that were included in the papers and written in widely used programming languages, such as R, Python and Perl.

The Mozilla engineers have discussed their findings with the papers' authors, who can now choose what, if anything, to do with the markups — including whether to permit disclosure of the results. Those findings will not affect the status of their publications, says Marian Petre, a computer scientist at the Open University in Milton Keynes, UK, who will debrief the reviewers and authors. Thaney expects to release a preliminary report on the project within the next few weeks.

Computational biologists are betting that the engineers will have found much to criticize in the scientific programming, but will also have learnt from the project. They may have been forced to brush up on their biology, lest they misunderstood the scientific objective of the code they were examining, Brown says.

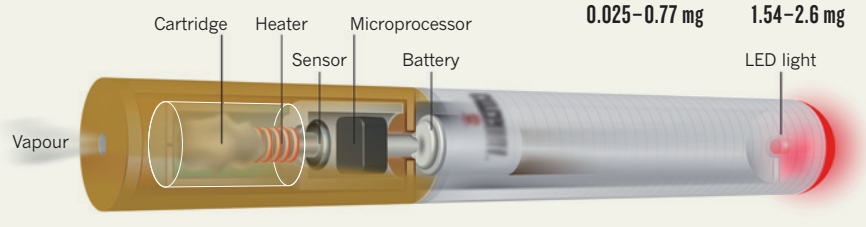
Theo Bloom, editorial director for biology at non-profit publisher PLoS, shares that expectation, but says such reviews may still be useful, even if the Mozilla reviewers lack biological expertise. Yet that would prompt another question: how can journals conduct this type of review in a sustainable way?

The time and skill involved may justify paying reviewers, just as statistical reviewers of large clinical trials are paid. But researchers say that having software reviewers looking over their shoulder might backfire. "One worry I have is that, with reviews like this, scientists will be even more discouraged from publishing their code," says biostatistician Roger Peng at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. "We need to get more code out there, not improve how it looks." ■

"One worry is that scientists will be even more discouraged from publishing their code."

SMOKE WITHOUT FIRE

Most electronic cigarettes use a battery-operated element to vaporize a nicotine solution, which is then inhaled. Doses of nicotine (right) and toxic chemicals (bottom) from e-cigarettes are lower than from conventional ones. In this comparison, 15 puffs on an e-cigarette are assumed to be equivalent to smoking a conventional cigarette.



COMPARISON OF TOXINS IN CONVENTIONAL AND ELECTRONIC CIGARETTES

Toxic compound	Conventional cigarette (μg in mainstream smoke)	Electronic cigarette (μg per 15 puffs)	Average ratio (conventional vs electronic)
Formaldehyde	1.6–52	0.20–5.61	9
NNN	0.005–0.19	0.00008–0.00043	380
NNK	0.012–0.11	0.00011–0.00283	40

NNN, *N*-nitrosornicotine; NNK, NNN and 4-(methylnitrosoamino)-1-(3-pyridyl)-1-butanone.

POLICY

Regulation stacks up for e-cigarettes

Devices may be the ‘healthy’ future of smoking — or a menace.

BY DANIEL CRESSEY

Electronic cigarettes could save the lives of millions of smokers, or they could set millions of non-smokers on the path to nicotine addiction, revolutionizing the tobacco industry into the bargain. So the question on the lips of health experts, policy-makers and consumers alike is, are the devices a health problem that needs tight regulation, or a welcome aid to smokers trying to quit?

In less than a decade since their first appearance, electronic or e-cigarettes have become a multibillion-dollar industry. Although there are scores of different products, most operate on the same principle: a heating element vaporizes a liquid containing nicotine, which can then be inhaled as ‘smoke’ (see ‘Smoke without fire’).

However, they are not without their dangers, and, as the number of users in the United States alone reaches an estimated 2.5 million, regulatory bodies have begun to take an interest. In October, the US Food and Drug Administration (FDA) is expected to issue a rule that affirms it has the authority to regulate e-cigarettes, overriding a previous court decision that e-cigarettes could not be controlled as medical devices. The European Union is also

overhauling its regulation of tobacco with a massive piece of legislation that, as currently drafted, will regulate most e-cigarettes as medical devices. A vote on this legislation is due in the European Parliament on 8 October.

But because little research has been done on the effects of e-cigarettes, such moves lack a solid scientific grounding. It is generally accepted that the devices are safer than conventional cigarettes, although studies by the FDA and Health New Zealand, a research consultancy based in Christchurch, have shown that some brands contain carcinogens and other toxic chemicals, including diethylene glycol and *N*-nitrosamines (A. D. Flouris and D. N. Oikonomou *Br. Med. J.* **340**, c311; 2010).

If e-cigarettes are used in moderation, the nicotine doses they provide may be lower than those attained from smoking cigarettes. But although the devices are smoke-free, nicotine itself causes high blood pressure and palpitations, and is highly addictive. Little is known about the long-term effects of e-cigarette vapour.

Some experts think e-cigarettes are a saviour. “They may kill smoking as we know it,” says Peter Hajek, director of the Tobacco Dependence Research Unit at Barts and the London School of Medicine and Dentistry.

“That’s the biggest hope we have of ending the tobacco epidemic.”

But as big tobacco companies have piled into a market worth more than US\$2 billion worldwide, regulators have failed to keep up, in part because the chemicals in e-cigarettes vary so widely. Some countries, such as Norway and Brazil, have banned the products. But in the United States, e-cigarettes are currently regulated only if they are marketed as quitting aids. The United Kingdom has said it will regulate them as medicines — meaning they will have to meet strict quality standards — but its regulator, the Medicines and Healthcare Products Regulatory Agency, is holding fire until the new European rules are in place.

The decisions that regulators make will shape not just the future of the industry but also the public-health response — and scientists both for and against e-cigarettes have waded into the debate while regulation is still up in the air.

“Right now, electronic cigarettes are the triumph of wishful thinking over data,” says Stanton Glantz, a tobacco-control researcher at the University of California, San Francisco, who thinks that the products should be regulated. He points to a report released earlier this month by the US Centers for Disease Control and Prevention in Atlanta, Georgia, that shows some children who have never smoked cigarettes are using e-cigarettes, suggesting that the devices may be a gateway product. And he notes that several surveys have reported high levels of smokers using both cigarettes and e-cigarettes, indicating that the products are being used to sustain nicotine addiction. The use of vapour flavourings, such as vanilla, could also be seen as an attempt to prolong use and appeal to younger consumers.

Other scientists, such as Hajek, say that regulating e-cigarettes as medical devices would be a disaster. He believes that the cost of complying with rules for medical devices would allow big tobacco companies to dominate the nascent e-cigarette industry, squeezing out innovative new products.

“To overregulate now could threaten the existence of e-cigarettes and cut down the options for people who want to quit,” agrees Christopher Bullen of the National Institute for Health Innovation at the University of Auckland in New Zealand. He was the lead author on a study published this month showing that e-cigarettes were as effective as nicotine patches in helping smokers to quit (*C. Bullen et al. Lancet* <http://doi.org/nq8>; 2013).

Vaughan Rees, a tobacco researcher at Harvard School of Public Health in Boston, Massachusetts, thinks that e-cigarettes need to improve before they can replace cigarettes — and that, for now, they should be regulated as tobacco products. Although they do present an opportunity to improve public health, he adds, care needs to be taken to ensure that they don’t flourish alongside conventional cigarettes. “Then we’ve got a double problem,” he says. ■

SOURCE: M. L. GONIEWICZ ET AL. NICOTINE TOB. RES. **15**, 158–166 (2012); M. L. GONIEWICZ ET AL. TOB. CONTROL <http://doi.org/nxy> (2013)

PHYSICS

Interface superconductivity found in single crystal

Iron-based compound revives search for room-temperature superconductors.

BY EUGENIE SAMUEL REICH

It is one of the more counter-intuitive stunts in the repertoire of scientists who study superconductivity. Stack two materials, even insulators, on top of one another and *voilà*: a dramatic drop in electrical resistance can emerge between them. The negligible resistance of superconducting materials, sought for applications such as highly efficient electric grids, typically emerges at temperatures close to absolute zero — and only in certain materials. But physicists have shown that the stacking effect, called interface superconductivity, can occur at temperatures up to almost 80 kelvin.

Now, a researcher says that interface superconductivity can exist in a single crystal of an iron-based superconductor. The discovery will rejuvenate work on a family of compounds that first showed high-temperature superconductivity in 2008, but in recent years had seemed to disappoint.

“This is something conceptually very new,” says Anvar Zakhidov, a physicist at the University of Texas at Dallas. “It raises fundamental questions that people will jump on.”

The provocative result comes from a superconducting pioneer: Paul Chu of the University of Houston in Texas. In the 1980s, he helped to discover the first high-temperature superconductors, a family of copper-containing compounds (see ‘All in the family’). In a paper posted on the arXiv preprint server on 30 August¹, Chu and his colleagues now report their work on an iron-based crystal that generally superconducts at 30 kelvin — but that contains small regions that superconduct at 49 kelvin. Chu proposes that the effect is the result of many interfaces forming tens of nanometres apart within these regions (see ‘At the interface’).

If that is the case, it would be the first time



Crystals of an iron compound may contain regions that superconduct at higher temperatures than the bulk.

that interface superconductivity has been detected in a stable, stand-alone crystal, instead of at the interface between separate materials that often deteriorate on heating.

Physicists searching for a material that superconducts at room temperature (290 kelvin or more) have been disappointed before: almost three decades of work with the copper compounds have failed to produce one. After iron-based superconductors burst onto the scene in 2008, chemists began tweaking them, and quickly brought superconducting temperatures above 50 kelvin. But then progress plateaued.

That changed in February last year², when physicists at Tsinghua University in Beijing

reported interface superconductivity at about 80 kelvin — above the temperature of liquid nitrogen. They had layered iron selenide, the simplest iron-based superconductor, with strontium titanate, a well-known insulator. “This was stunning — the fact that we are able to construct new superconductors artificially,” says Jean-Marc Triscone, a condensed-matter physicist at the University of Geneva in Switzerland.

But Chu’s result goes further. He has reported the effect in a single crystal of a calcium-iron-arsenide compound doped with rare-earth elements, and suggests that the interfaces sit between regions that are enriched and depleted with arsenic.

JOHNPIERRE PAGLIONE



**MORE
ONLINE**

INTERVIEW



President of Russian Academy of Sciences hits out at political meddling in research go.nature.com/16w6re

MORE NEWS

- Relic radiation suggests the Universe may not be flat go.nature.com/szagdh
- Earth will leave the Sun’s ‘habitable zone’ in 1.75 billion years go.nature.com/rqpnqn
- Rover sees no sign of methane in Martian atmosphere go.nature.com/jdrgxk

NATURE PODCAST



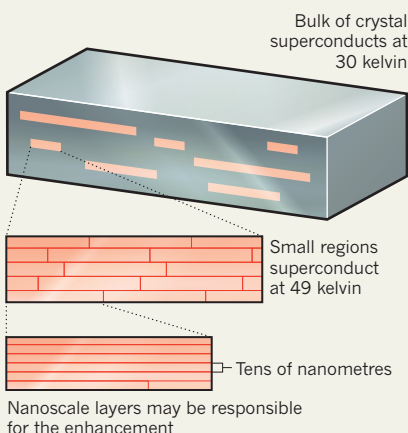
A carbon-nanotube computer, saving species by tweaking their genes, and early Earth’s similarity to Io nature.com/nature/podcast

If correct, the result would show that interface superconductivity is consistent with a mechanism first advanced in 1973, which helps to explain low-temperature superconductivity. Proposed by US physicist John Bardeen and his colleagues, it posits that superconducting electron pairs are held together by an exciton, a particle produced by the scattering of electrons off a boundary. This does not explain all high-temperature superconductivity, but exploring it further might shed light on mechanisms that do.

Superconductivity is a field with a long history of mistaken claims and false hopes, and so it is unsurprising that Chu's result is provoking controversy. Critics say that it is difficult to measure superconductivity along narrow interfaces, let alone inside a crystal. And Johnpierre Paglione, a condensed-matter physicist at the University of Maryland in College Park, is dubious for a different reason. In 2012, he and his team reported³ a higher-than-expected superconducting temperature for regions within the same material. But in a paper posted on arXiv on 13 September⁴, the researchers say that what they saw is a bulk property of these regions,

AT THE INTERFACE

Nanoscale layering in an iron-based crystal could enhance its superconductivity.



caused by the doping with rare-earth elements, rather than an interface effect.

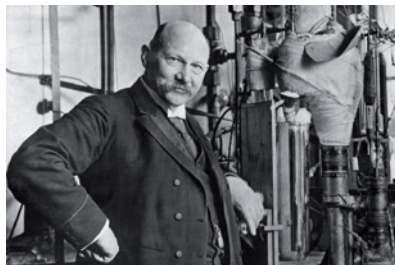
If the superconductivity was related to interfaces, Paglione says, he would not expect it to disappear when the doping is reduced — which it does. Instead, Paglione suggests that strain in certain regions of the crystal enhances superconductivity. Chu responds that doping above a certain level helps to create interfaces by pushing arsenic atoms out of the structure.

Ivan Bozovic, a physicist at Brookhaven National Laboratory in Upton, New York, who in 2008 created one of the first interfaced superconductors, raises another possibility. He says that the superconductivity might not be an interface effect; instead, it could be attributable to a structural change inherent to the arsenic-depleted regions. He would like to see whether Chu can isolate these regions to study them further — and perhaps use them to create other interfaces. “We’ve been looking for new superconductors for 100 years and interface superconductivity is a new trick,” says Bozovic, “so there’s a lot of room for new combinations.”■

1. Wei, F. Y. *et al.* Preprint at <http://arxiv.org/abs/1309.0034> (2013).
2. Wang, Q.-Y. *et al.* *Chin. Phys. Lett.* **29**, 037402 (2012).
3. Saha, S. R. *et al.* *Phys. Rev. B* **85**, 024525 (2012).
4. Saha, S. R. *et al.* Preprint at <http://arxiv.org/abs/1309.3576> (2013).

ALL IN THE FAMILY

Superconducting temperatures have risen as new classes of material have been found.



1911 4.2 kelvin: Mercury. Superconductivity first discovered by Heike Kamerlingh Onnes (pictured).

1986 30 kelvin: Barium lanthanum copper oxide. First copper-based high-temperature superconductor.

1993 133 kelvin: Mercury barium calcium copper oxide. Record high-temperature superconductor.

2008 26 kelvin: Lanthanum iron arsenide. First iron-based high-temperature superconductor.

2008 50 kelvin: Layered insulating and metallic copper oxide. One of the first interface-enhanced superconductors.

2013 49 kelvin: Calcium iron arsenide doped with rare-earth elements. Claimed single-crystal interface enhanced superconductor.

CORRECTIONS

The scale on the graph showing German science spending in the News story ‘Germany hits science high’ (*Nature* **501**, 289–290; 2013) had an incorrect scale. It should have run from €0 to €15 billion not €0–1.5 billion. And the News story ‘Grass gets greener’ (*Nature* **501**, 291; 2013) incorrectly identified the grain that produced high yields when grown in a field that previously hosted *Brachiaria* grasses. It was maize (corn), not wheat.

THE START-UP ENGINE

Third Rock Ventures made its name by placing big bets on the biotechnology companies it launched. Now, everyone is waiting for the pay-off.

BY HEIDI LEDFORD



Bioengineer Mikhail Shapiro got a rude shock one day when he arrived for work at Third Rock Ventures, then a brand-new venture-capital firm headed by a handful of biotech elites. Only three weeks into his internship, Shapiro found a notice on the door: “Closed for business.” Inside, ‘For sale’ signs hung on desks, equipment, everything — even the office’s giant gumball machine. The company had folded, a note explained, because it could not raise enough money.

Kevin Starr, a partner at Third Rock, still beams with pride over that 2007 prank, which he and his confederates had filmed to capture Shapiro’s reaction. “You could tell Mikhail was thinking, ‘I knew that was going to happen to these guys!’” he recalls.

Few would fault Shapiro, now a professor at the California Institute of Technology in Pasadena, for his credulity. By 2007, the technology bubble of the early 2000s had burst, and investors were baulking at the long timelines and high failure rates involved in getting biotechnology products to the market. People laughed, says Starr, when he and Third Rock’s other founders told them that the company wanted to raise US\$378 million to create an investment fund to build biotech companies from scratch. “They advised us to aim for about a tenth of that.”

But Third Rock, based in Boston, Massachusetts, did raise its initial fund, and it has not slowed down since. The company has brought in \$1.3 billion and invested in more than 30 young companies, many based on cutting-edge research in fields such as cancer epigenetics, gene therapy and medical diagnostics (see ‘Due diligence’).

Products are only just starting to trickle out into clinical testing, but this year brought several signs that the firm has bet well. In January, Third Rock sold off Lotus Tissue Repair — a tissue-engineering company with an experimental therapy for a devastating rare disease that weakens skin. The deal could garner a 20-fold return for Third Rock if Lotus meets certain milestones. In March, Third Rock’s third round of funding — \$516 million to launch up to 16 more companies — had so many aspiring investors that the firm had to turn some away. And this summer, two of Third Rock’s companies went public, their share prices soaring the moment they hit the market. As *Nature* went to press, a third firm — cancer diagnostics company Foundation Medicine in Cambridge, Massachusetts — was preparing to follow suit.

“For a long time, people said investing in these early-stage companies was not a great idea,” says Robert Langer, a bioengineer at the Massachusetts Institute of Technology (MIT) in Cambridge who has spun off dozens of companies from his research (see *Nature* 458, 22–24; 2009). “Third Rock has taken that risk and I think it’s paying off.”

LAID-BACK BIOTECH

Since 2007, Third Rock has expanded its offices on Boston’s trendy Newbury Street — a neighbourhood filled with high-end boutiques and cafes. On a flaming day this summer, Starr sits in his office arrayed in silver jewellery, camouflage shorts and a green T-shirt that reads “Beach Punk 1982”. A standard business shirt bides its time on a hanger behind the door.

Starr’s laid-back style has found lots of attention in the business press, and it serves as a reminder that he does not have to be here. In 2003, he left a post as chief operating officer of Millennium Pharmaceuticals, a Cambridge-based biotech powerhouse that had just launched the blockbuster cancer drug Velcade (bortezomib). Millennium founder Mark Levin retired some time after Starr, and the two did the usual

things that young retirees with plenty of money do — travelling the world and producing independent films and Broadway shows. In 2006, Starr says, during an annual pilgrimage to the golf courses and blackjack tables of Las Vegas, Nevada, Levin turned to him and said, “Hey

Kev, why don’t we just go do something again?”

Venture capital has a pivotal role in transforming science into medical advances, supporting companies during the long, lean, research-intensive years before they have any hope of turning a profit. In the United States, biotech soaks up billions of dollars in venture capital each year, second only to the software industry. In the mid-2000s, infusions into fledgling companies made up just a tiny fraction of that investment. Most of the money was going to established companies, often with products already in clinical testing. But the pharmaceutical industry was tightening internal research budgets and looking to small biotechnology firms for new medicines.

Amid that changing landscape, Starr and Levin saw an opportunity. There would be demand for innovative biotechnology companies, yet few venture capitalists were in a position to fill it. Through a series of meetings at Starbucks, Levin and Starr assembled a skeleton crew of biotech nobility and mapped out their ideal venture-capital firm.

STANDING OUT

Levin, Starr and Bob Tepper, former head of research and development at Millennium, wanted to do things differently from typical venture capitalists, who sift through ideas and business proposals from external researchers, help to set up a company and then hand over control to a newly recruited executive team. Starr says that he and his co-founders wanted to recreate some of the magic they had felt at Millennium, carrying over its ‘anything is possible’ mantra. They would hire only the best people, even if that meant interviewing candidates for months. And, rather than relying on proposals from the outside, they would focus on the hottest science, mostly investing in companies conceived by Third Rock’s team. “Last year we saw 982 outside plans,” says Starr. “We invested in zero.”

All venture capitalists need to understand the science behind their investments, but Shapiro, who has since worked with other venture-capital firms, says that Third Rock is unique in how far its members personally immerse themselves in the details. “It’s a bunch of nerds,” he says. “You’re in a commercial setting, but the rigour of the science was as high as it was at MIT or Caltech.” Of the more than 40 employees now at Third Rock, only Levin, a chemical engineer by training, had worked

in venture capital before. The rest had trained in the trenches as scientists, physicians and biotech business leaders. “They have decades of real, hands-on experience,” says Michelle Dipp, a venture capitalist at the Longwood Fund in Boston. “It’s an incredibly talented team.”

Third Rock also takes its time handing over the reins of its companies to outside executives; it often waits 18 months or longer. That is important for luring top talent, says Langer. “A lot of good chief executives are not willing to take the risk with a new company,” he says. “With Third Rock, rather than getting the company when it’s a newborn baby, a new executive

is getting a pretty active 2-year-old.”

Finding newborns to raise means exploring promising ideas, something that Third Rock spends about one-third of its time doing. Those that pass muster get up to \$2 million and must go through a rigorous and lengthy screening process that employees refer to as the ‘Third Rock Ultra Killer Criteria’ (TRUKK). Independent labs must be able to replicate key findings and find no warning signs of toxicity for drug candidates.

Third Rock also circulates the project idea to contacts at large pharmaceutical firms. If those companies have internal scientists working on the same project, Third Rock generally will not try to compete. Or if pharmaceutical insiders say that they like the idea but would not invest in it without seeing data from late-stage clinical trials, the project is scrapped. For all the talk of how anything is possible, Third Rock is ruthlessly practical.

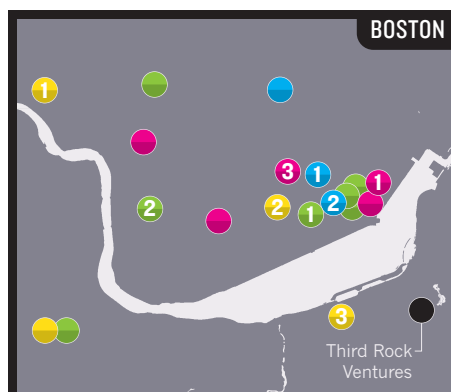
To meet the TRUKK, a project must be no more than about three years away from clinical testing — a brutal necessity of the ten-year

**“YOU’RE IN A
COMMERCIAL SETTING,
BUT THE RIGOUR OF
THE SCIENCE WAS AS
HIGH AS IT WAS AT MIT
OR CALTECH.”**

Kevin Starr (right) and Mark Levin founded one of the hottest venture-capital firms in biotechnology.

DANA SMITH

DUE DILIGENCE *Third Rock Ventures has a portfolio of more than 30 companies that are developing drugs and therapies in several fields.*

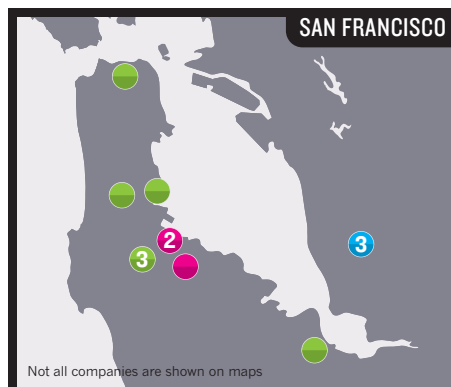


CANCER

- 1. Constellation Pharmaceuticals**
Targets proteins that regulate the genome epigenetically, such as the methyltransferase EZH2.
- 2. Igenica**
Aims to identify effective antibody treatments for cancer.
- 3. Foundation Medicine**
Develops tools to profile cancers genetically and better match patients to therapies.

METABOLIC DISEASES

- 1. Ember Therapeutics**
Develops diabetes and obesity drugs based on activating brown-fat stores in the body, among other mechanisms.
- 2. Zafgen**
Targets the enzyme methionine aminopeptidase 2, which regulates how the body metabolizes fat.
- 3. Rhythm**
Develops peptides to tackle obesity by regulating appetite-stimulating hormones.



DEVICES

- 1. NinePoint Medical**
Develops systems to perform real-time pathology on patient tissues without the need for biopsy.
- 2. Seventh Sense Biosystems**
Works on devices to track patient health by monitoring changes in the skin.
- 3. Cibirom**
Is testing a catheter-based approach to control symptoms of cardiovascular diseases, diabetes and renal failure by modulating the sympathetic nervous system.

OTHER

- 1. Edimer**
Works on X-linked hypohidrotic ectodermal dysplasia, which is associated with respiratory problems and a lack of sweat glands.
- 2. bluebird bio**
Develops gene-therapy approaches to treat conditions such as sickle-cell anaemia and childhood adrenoleukodystrophy.
- 3. Afferent Pharmaceuticals**
Targets a nerve receptor called P2X3 to treat chronic pain conditions associated with inflammation and injury.

funding cycles of venture capital, says Starr. This can mean painful decisions. A few years ago, the team evaluated the therapeutic promise of a class of gene regulators called long non-coding RNAs. The field is hot and the team loved its potential, but it was not close enough to the clinic to serve as the basis for a Third Rock company.

Although draconian on paper, Third Rock has bent its rules for some early investments. Agios Pharmaceuticals in Cambridge develops drugs to target metabolic changes that fuel tumours. It was among the first companies Third Rock backed, in 2007, but it was August 2013 before the firm began clinical trials on its lead compound — a drug to combat tumours that contain mutations in the metabolic gene *IDH2*.

Agios's chief executive, David Schenkein, says that the company's early drug leads did not work, forcing scientists to develop new candidates from scratch, but he has felt no pressure from Third Rock about the revised timeline. Starr says that Third Rock factors a few setbacks like this into its calculations. And when Agios went public in July, investors showed that the delay did not concern them either: Agios made \$106 million and its stock rose 56% on the first day of trading.

DANGEROUS PRECEDENT

Some will watch companies such as Agios carefully, mindful of the history of Millennium Pharmaceuticals. Millennium was founded in 1993, intended — like many of Third Rock's companies — to be a 'product engine': a company based on a single scientific premise or technology that could generate multiple therapies. In Millennium's case, that premise was personalized medicine grounded on emerging human genomics data. That mission failed and, as with other firms of the era, Millennium's internal research programme foundered, burning up cash in the process. But the company did use its expertise to identify promising drug candidates from outside firms, bringing them into Millennium for the final stages of development and marketing.

NATURE.COM
Is the biotech bubble ready to burst? See: go.nature.com/hxtggd

Long-term investors benefited: in 2008, the Japanese pharmaceutical company Takeda, based in Osaka, bought Millennium for \$8.8 billion.

There may be debate about whether Millennium was a scientific success, but there is no doubt that the company had a tremendous impact on the booming Boston–Cambridge biotechnology industry. Former employees sit at or close to the helms of biotech companies across the region. They are often quick to credit Levin and the culture he created at Millennium. "We came in as scientists," says Rosana Kapeller, once director of molecular and cellular biology at Millennium and now chief scientific officer of Nimbus Discovery, a pharmaceutical company in Cambridge. "We left as entrepreneurs."

BRIGHT PROMISE

That environment allowed Millennium to succeed even after its initial mission failed, says Starr. "We built a company with the right formula and culture to be a successful long-term company," he says, "as opposed to a company that just stuck with one model and ran right into the wall."

Agios may be echoing Millennium in that respect. In addition to developing new compounds, it last year broadened its mission to include genetic diseases of metabolism. The firm has developed a compound that will enter clinical testing next year in patients with pyruvate kinase deficiency, a rare metabolic condition that causes severe anaemia.

Years will pass before Third Rock's success — whether measured in medical breakthroughs or in cash returns to investors — can be assessed. But people familiar with the team are certain of one thing: the pranks will continue. Asked to confirm the details of the joke that Starr played on him in 2007, Shapiro was confused. "Which one?" he asked.

"It's a really high-powered group of people with an incredible track record of achievement," he says. "But it's also a group of people who don't take themselves too seriously." ■

Heidi Ledford reports for Nature from Cambridge, Massachusetts.

OCEANOGRAPHY'S BILLION-

A mammoth undersea US project will soon start streaming data to researchers. But some wonder whether the system is worth its high price.

BY ALEXANDRA WITZE

On a sunny day in July, it takes 90 minutes for the R/V *Thomas G. Thompson* to traverse the locks connecting Seattle's inland waters to Puget Sound. On deck, John Delaney looks impatiently out to sea. As an oceanographer at the University of Washington in Seattle, he has made this trip many times to explore beneath the Pacific's waves. But Delaney hopes that this seven-week expedition will be the beginning of the end for these time-consuming journeys. "We don't want to be ship-bound," he sighs. Instead, he is spending money — a lot of money — to open a permanent window onto the sea floor.

Delaney is the architect behind a 925-kilometre network of fibre-optic cable and instruments being installed on the seabed off the coast of Washington and Oregon. If all goes according to plan, these will stream real-time data back to shore by 2015, delivering some of the first live video footage of an underwater volcano erupting, hydrothermal vents growing and clouds of microbes billowing from the sea floor. The cabled network is a key part of the massive US Ocean Observatories Initiative (OOI), which aims to create a flood of continuous information from select sites.

Oceanographers have long relied on brief glimpses of data from single research cruises or isolated buoys or moorings. The OOI, and Delaney, aim to exchange those flashes of insight for a constant spotlight. "The goal is to launch an era of scientific discovery," Delaney says, thumping his fist on the ship's deck rail. "This is a game-changer."

Many US oceanographers have not yet considered just how the OOI's broad scope and potential might affect their research. But some who have been watching its development closely warn that the project is an expensive gamble. Construction costs will run to US\$386 million, and the programme will then consume about \$55 million per year for operations and maintenance. By the end of its planned 25-year lifetime, the OOI will have cost nearly \$1.8 billion — an unprecedented price tag in oceanography.

The running costs will eat up about one-sixth of the annual budget for ocean sciences at the US National Science Foundation (NSF), and that proportion could increase. "That money is being pulled right out of what could otherwise be allocated for peer-reviewed science," says Charles Eriksen, an oceanographer at the University of Washington who is not involved with the project. Critics also complain that the OOI sites cover only a fraction of a per cent of the world's oceans.

Such objections hold little water with Delaney. "People say cables are expensive," he says. "Well, ships are expensive too." The hour and a half it takes to traverse Seattle's locks costs the University of Washington roughly \$4,000.

TELECOM SPIN-OFF

The OOI sprouted from a germ of an idea planted by Delaney and Alan Chave, an ocean scientist now at the Woods Hole Oceanographic Institution (WHOI) in Massachusetts. It was the early 1990s, and Delaney was frustrated with getting only enough ship time to visit the sea floor for a day or two every couple of years. Chave had been working for the US telecommunications giant AT&T, and he suggested hooking a piece of old telephone cable up to instruments on the seabed. The fibre-optic

cable would provide electricity and stream data back to shore. "This is transformational," Delaney thought.

The idea of a permanent oceanographic observatory slowly gathered steam at the NSF, and by 2007 the agency had decided to invest some \$330 million in the concept. The problem was working out exactly what to build. To qualify as a national facility and justify the cost, the observatory needed to expand beyond sea-floor cables to include instruments that could plumb the full depth of the water column. OOI proponents initially dreamed up an observatory on steroids: multiple cabled arrays along with more than a dozen coastal and deep-water sites. Not surprisingly, the final project design cut back on most areas. "You have to build

what you can afford," says Deborah Kelley, a marine geologist at the University of Washington and an OOI project scientist.

After input from hundreds of researchers, the project team settled on three main components (see 'A mega-ocean observatory'). The first, and most ambitious, is the fibre-optic network southwest of Seattle, most of which has now been laid. This will connect dozens of sea-floor instruments across the Juan de Fuca tectonic plate, which slides under North America and drives seismic activity along the west coast from northern California to Vancouver Island

in Canada. The instruments will focus on an active underwater volcano called Axial Seamount, and a formation called Hydrate Ridge, where methane vented from the sea floor feeds a unique ecosystem.

The second component involves laying an array of moorings to support instrumentation off the east and west coasts of the United States. In each array, automated profilers will shuttle up and down cables measuring chlorophyll, oxygen and other factors from the sea floor to the surface. Six gliders will rove between moorings to make similar measurements. The third part of the project will use moorings and gliders to monitor four deep-water sites in the far north and south. These six sites will be run by WHOI, the Scripps Institution of Oceanography in La Jolla, California, and Oregon State University in Corvallis.

Together, these stations will marshal the forces of about 760 sensors, of 47 different designs, to collect data on variables ranging from water temperature, salinity and density to acidity, carbon dioxide and oxygen levels. "One of the most transformational things is how interdisciplinary it will be," says Kelley.

Elsewhere, similarly ambitious oceanographic observatories are already in use. Japan has two dense sea-floor observatories — DONET and DONET2 — with a focus on earthquake and tsunami studies. And many nations operate networks of buoys: the international Argo project has an array of more than 3,000 floats. The OOI will collect a broader selection of data than those efforts. In Canada, a cousin to the OOI has been up and running since late 2009. That project, called NEPTUNE Canada, involves 800 kilometres of fibre-optic cable laid on the northern part of the Juan de Fuca plate. The Canadian government has spent Can\$200 million (US\$194 million) on sea-floor observatories. The OOI's cabled observatory will be very similar, but because it is

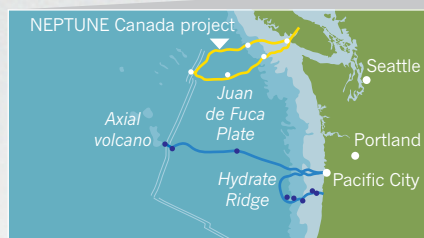
**"THE GOAL IS TO
LAUNCH AN ERA
OF SCIENTIFIC
DISCOVERY.
THIS IS A GAME-
CHANGER."**

PHOTO: ALLISON FUNDIS, UNIV. WASHINGTON; GRAPHIC: NIK SPENCER/NATURE

DOLLAR BABY

A MEGA-OCEAN OBSERVATORY

The US Ocean Observatories Initiative includes one sea-floor cable observatory (called Regional Scale Nodes), four deep-sea sites, and two coastal arrays of instruments; but together, this covers only a fraction of a per cent of the world's oceans.



- Deep-sea site
- Coastal array
- ★ Science highlight

REGIONAL SCALE NODES

800–2,900 m water depth
925 km of cable
7 sea-floor terminals

- ★ Earthquakes, volcanoes, methane hydrates

STATION PAPA

4,250 m water depth
1 full-depth profiler
2 subsurface moorings
3 gliders

- ★ Historic site with records dating back to 1949

ENDURANCE ARRAY

25–600 m water depth
2 mooring lines
6 gliders

- ★ Coastal upwelling important for fisheries

IRMINGER SEA

2,800 m water depth
1 surface mooring
1 full-depth profiler
2 subsurface moorings
3 gliders

- ★ Productive fisheries area

PIONEER ARRAY

130 m water depth
10 moorings
6 gliders
3 autonomous underwater vehicles

- ★ Gulf Stream boundary currents; array can be moved to other sites in future

Atlantic Ocean

Pacific Ocean

ARGENTINE BASIN

5,200 m water depth
1 surface mooring
1 full-depth profiler
2 subsurface moorings
3 gliders

- ★ Closes a gap in the global seismic network

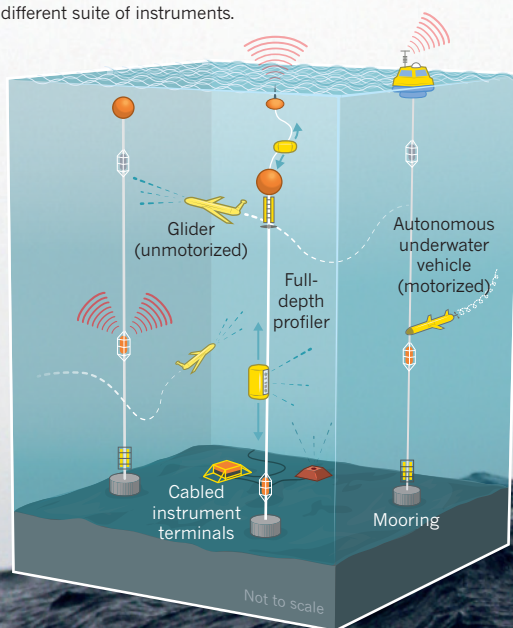
SOUTHERN OCEAN

4,800 m water depth
1 surface mooring
1 full-depth profiler
2 subsurface moorings
3 gliders

- ★ Region absorbs significant amount of carbon dioxide.

MARINE TOOLBOX

The project's many instruments, including those shown here, incorporate a total of about 760 sensors. Each research site includes a different suite of instruments.



getting a later start, it will have newer designs of sensors and moorings, says Kate Moran, president of Ocean Networks Canada, which oversees NEPTUNE Canada.

It is these sensors that Delaney is dreaming of as the R/V *Thompson* reaches the Axial volcano in July. When the captain pulls the ship to a halt above the underwater mountain, engineers manoeuvre a boxy yellow submersible called ROPOS down into the water. From 1,500 metres below the ocean surface, it sends back a murky video feed of an eight-metre-high mineral chimney, its sides festooned with microbial growth, its top spouting shimmering hot water.

SNAPSHOT SCIENCE

Visiting Axial like this costs almost \$70,000 a day and provides only a snapshot of the volcano's behaviour. Oceanographers have, in the past, dropped off seismometers and hydrophones at Axial and retrieved them later, allowing them to study an eruption after the fact. And, in 2011, an expedition led by the US National Oceanic and Atmospheric Administration happened to arrive just months after the volcano erupted, providing scientists with a fortuitous opportunity to study fresh lava flows. But sightings of underwater eruptions are rare, even though this is the most common type of volcanism on Earth.

Axial is expected to erupt again within a decade and OOI researchers plan to catch it in the act. A bottom-pressure tiltmeter will measure gradual changes that could indicate that the volcano is inflating, and a cutting-edge mass spectrometer will sniff the water for hints of magma rising from below. When an eruption seems imminent, a fleet of gliders could be deployed to study chemicals in the water column. A high-definition video camera, installed this summer, will watch a hydrothermal vent on Axial's flanks. All this should yield details about how magma reaches the volcano's summit, how that relates to seismic activity, and how organisms living in this extreme environment deal with an eruption. "There are so many questions we don't have answers to," says Kelley.

Delaney says that the \$126-million cabled observatory is the only way to spy on Axial properly. The OOI's investment in other parts of the ocean will also gather data that cannot be obtained using existing equipment, says Tim Cowles, programme director at the Consortium for Ocean Leadership, the organization in Washington DC that is overseeing the OOI.

The four high-latitude deep-sea moorings, for example, will help to study the exchange between air and water where powerful winds whip the sea surface into a froth. Weather forecasters and climate modellers need better information on how energy and gas move between the deep ocean, the sea surface and the atmosphere. But because oceanographic equipment tends to take a beating in fierce weather, measurements are few and far between at high latitudes. A team led by Uwe Send at Scripps deployed 57 instruments at a site in the Gulf of Alaska in July, making it the first deep-sea OOI site to be completed.

The coastal arrays, meanwhile, should be useful for studies of algal blooms. Gradual changes in near-surface temperatures can drastically affect how phytoplankton blooms form in the spring, but catching such changes in the act is tricky, says Kendra Daly, a biological oceanographer and OOI project scientist at the University of South Florida in St Petersburg. "We have no ability to predict exactly when that will happen, so it's hard to get ship time to be out there right when the bloom starts and ends," she says. Measuring nutrient concentrations and primary productivity on the spot will allow researchers to better quantify how much carbon dioxide is absorbed by blooms, which has implications for understanding how biological systems interact with climate, Daly says.

To even start to tackle these science questions, the OOI team will need to get its instruments into the water before a tight deadline of February 2015. After that, the money available to complete the job will dry up. "All those things that have to happen in the last six months of the project do make all of us nervous," says Cowles. The project's schedule has been frantic since 2009, when the OOI received an initial input of almost \$106 million from the government stimulus bill enacted in the

wake of the economic recession. That infusion of funds was welcome, but it left the team scrambling to get everything in place. "It kicked us out of the gate ahead of schedule, and we've been trying to catch up ever since," says Cowles.

At least one hurdle stands in the way of a timely completion. For now, the team led by the University of Washington can't connect the instruments it has deployed to the backbone power-and-data cables leading to the shore. The NSF requires that the company that made the cables, L-3 MariPro of Goleta, California, ensure that they are in proper working order before handing them over, and L-3 MariPro is running behind. Delaney's team had to shorten this summer's field season and bank some of its ship days for 2014, intending to connect the instruments then.

And there are other, bigger worries. One is the daunting cost of maintenance. OOI sites are slated to be serviced every one, three or five years — depending on the equipment — which will run up a hefty ship-time bill. And unforeseen glitches are bound to strike. NEPTUNE Canada has run into major technical issues, including the failure of kilometres-long segments of cable and instruments that stopped working after just a year on the sea floor.

BY THE END OF ITS 25-YEAR LIFETIME, THIS OBSERVATORY WILL HAVE COST NEARLY \$1.8 BILLION.

Ensuring data quality is also a concern. "That's the big worry in my mind," says Douglas Luther, an oceanographer at the University of Hawaii at Manoa and an OOI project scientist. The OOI's cyberinfrastructure team is developing automated algorithms to flag up any obvious problems — such as sensors that record temperatures hundreds of degrees above those of neighbouring devices — but there is currently no money to fund a big quality-control team.

Another question is how much demand there will be for the data. Not everyone will be able to abandon field trips in favour of using the OOI's instruments. Microbial oceanographer Julie Huber of the Marine Biological Laboratory in Woods Hole studies microbes living on and in Axial volcano. The OOI's cabled network isn't much use to her — so far there is no instrument she could plug in for her microbial monitoring. "It doesn't replace me having to go out there," she says.

OCEAN OUTREACH

The OOI's leaders could follow the example of their northern colleagues. Canada's undersea cabled networks, which cost Can\$16 million annually to run, have nearly 8,000 active data users per year, which is right around the level their funders wanted to see. In large part, that is because Ocean Networks Canada employs six staff scientists to reach out to researchers and educators with suggested ways to use the data. "Users just don't come — you have to work at it," says Moran. So far, the OOI has no formal outreach plan.

There are also fears that NSF programme managers will feel under pressure to fund projects that use OOI data, in order to validate its cost. "Mostly, we'll be paying a lot of money trying to make this hardware a success," says Russ Davis, an oceanographer retired from Scripps who helped to develop key floats for the Argo array. "If it isn't awfully wonderful, it's going to look bad for the NSF and be bad for science."

For Delaney, the OOI's potential outweighs such concerns. As the *Thompson* makes its way across the Pacific, leaving the underwater volcano behind, he muses about the interconnectedness of the oceans. "A single ship can only be in one place at one time," he says. "We need to be present in multiple places in multiple times."

That omnipresent capability is what the OOI is all about. And if it costs a lot of money, Delaney wants the research community to keep the sums involved in perspective: NASA, for comparison, spends billions each year. "Our investment in the ocean is way below our investment in outer space," he says. "But the return is much greater." ■ SEE EDITORIAL P.461

Alexandra Witze reports for Nature from Boulder, Colorado.

COMMENT

CONSERVATION Could genetic engineering preserve species? **p.485**

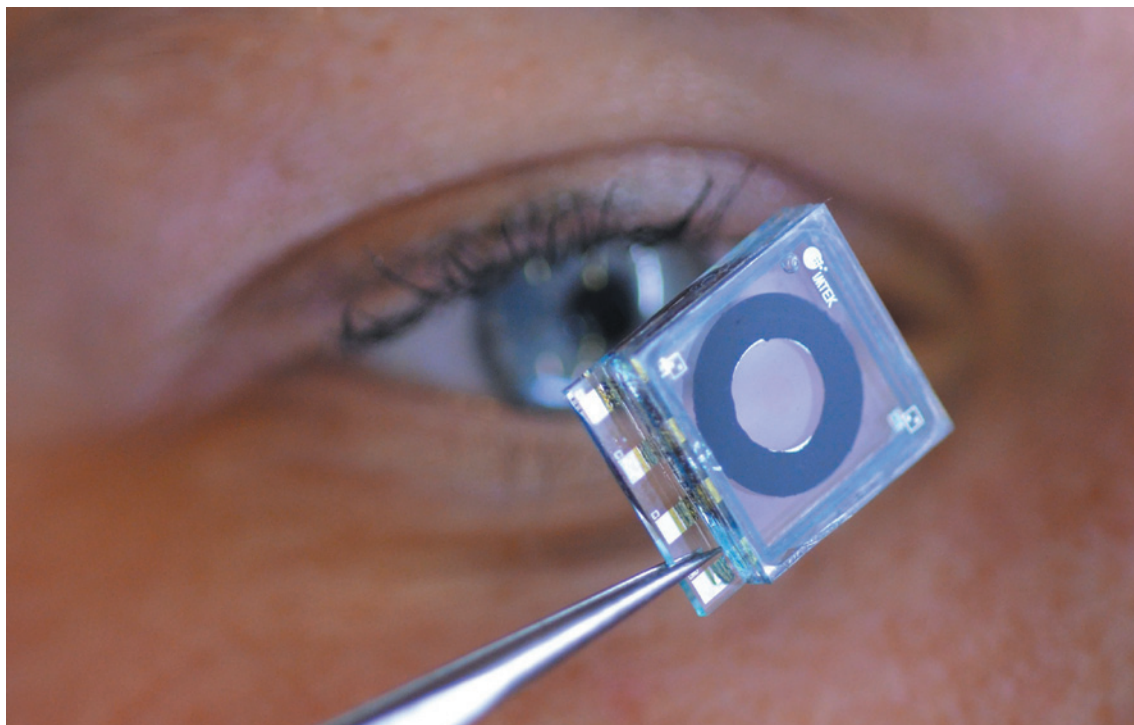
BIOGRAPHY The friendship of Nobel laureates Jacques Monod and Albert Camus **p.487**

HISTORY Britain's role in the race to build the first atomic bombs **p.488**



TECHNOLOGY Two takes on paper and punctuation **p.491**

LABORATORY FOR MICRO-OPTICS UNI. FREIBURG



Optofluidic chips must meet demanding industry specifications before making it to market.

Bridging the market gap

Physicists and engineers must do more than peddle ideas if their technologies are to translate effectively beyond the lab, says **Hans Zappe**.

Earlier this year, I witnessed a scout from a consumer electronics firm visiting a university lab to assess a prototype optical chip for potential investment. Holding the fluid-filled device between his fingers, and with an eye on the postdoc who had produced it, the visitor gently shook it. Nothing happened. He shook it more violently. Still no leaks. The postdoc smiled. The scout threw the chip to the floor. Coolly, the researcher picked it up, placed it under a microscope and showed the guest that it still worked perfectly. The prototype was surely well advanced.

The researcher's smile faded some weeks later, when a development contract from the electronics firm arrived in the post. The contract made it clear that the two parties had

disparate views on what they considered to be advanced. The industrial partner imagined paying for 100 fully characterized and functioning samples and instructions on how to make them — to be delivered as soon as possible. The university partner looked at the long list of specifications, ranging from sub-freezing storage temperatures to device lifetimes measured in years, and realized that much more effort would be required. Only then did the difficult development work

NATURE.COM
See *Scientific American's*
innovation special:
go.nature.com/chobc2

begin: setting out a detailed programme in which the performance of the chips could be tested, assured and delivered.

It is a familiar tale. Although translational research is established in biomedicine, yielding many commercial pharmaceuticals (see *Nature* 453, 830–831; 2008), technology translation is still a big problem in the physical, chemical and biological sciences, and in engineering (see 'Patent problem'). There can be a yawning gulf between an exciting result that is suitable for a paper in *Nature* or *Science*, and its realization in a form that allows a company to begin product development. For example, some advanced miniaturized optical-imaging systems for medical diagnostics have yet to find manufacturers, even five or more years after the concept has been proven.

The technology-translation gap exists ►

► for many reasons: cultural, institutional and technical. To bridge it, universities and companies must better understand each other's needs, motivations and limitations. The fastest way to close the gap is to spend more time together in each other's labs.

DIFFERING VIEWS

The cultural barriers to technology translation are ingrained in both camps. Companies would like universities to distribute the fruits of their research free of charge, because, after all, much of the work has been paid for with public money. Engineering departments, in their view, should focus on developing prototypes that can be manufactured on a large scale, and not 'waste' time with blue-sky research that has no obvious commercial value.

In many university departments, by contrast, industrial collaboration is regarded with suspicion. Academics shy away, fearful that companies will influence research directions and covet useful results.

In the United States, reluctance to partner with industry began to thaw with the passage in 1980 of the University and Small Business Patent Procedures Act, known as the Bayh-Dole Act. The act allowed government-funded educational institutions to retain the titles to their patents. Universities discovered a lucrative source of funding, and technology-transfer offices sprouted up like mushrooms on US campuses.

Other countries followed suit, but did not always see the same results. The equivalent law enacted in Japan in 1999 boosted the number of research-and-development projects between national universities and private industry from 56 in 1983 to 14,303 in 2008 (ref. 1). But in Germany, which already had a tradition of assigning patent rights to the inventing faculty member², the passing of similar legislation in 2002 led to stagnation in patent activity³, because some

researchers held back their patentable ideas.

To improve matters, universities that are keen to build up their patent portfolios need to provide more incentives for individual academic researchers to engage with the technology-transfer process, which can be tedious and expensive. Compared with publications, patents carry little weight in most academic evaluations. To establish a 'patent culture' on campus, policies, rules, rewards and ethos must support faculty involvement in business activities⁴.

Culture alone may explain why electrical engineers at Stanford University in California, for example, have 50% more corporate affiliations (253 in 2004) than their colleagues at the University of California, Berkeley (168 in 2004), even though the departments are similarly sized, both rank in the top few in the United States and both are close to Silicon Valley⁵. Whereas Stanford — the alma mater of the founders of Google, for instance — has a strong tradition of valuing such activities, the University of California system is more ambivalent.

Even when campus culture supports entrepreneurship, there is often a further hurdle: the unrealistic financial or intellectual-property expectations of a university's industrial liaison office. Inspired by a few industrial geese who did lay golden eggs, the lawyers in many university technology-transfer offices have scared off potential collaborators by demanding too much. Whether these offices foster or hinder fruitful collaboration is controversial⁶. In my experience, their utility varies strongly with the personalities and qualifications of their staff.

As in the opening tale, diverging technical expectations mar many industrial-academic relationships. 'University prototypes' and

'industrial prototypes' are distinct. University researchers love to pursue wild ideas and to artfully perform difficult experiments that, even if challenging to reproduce, need only work a few times to yield a high-profile paper that could advance their academic career. For the industrial researcher, a device has to work every time, with well-understood reliability, reproducibility and lifetime.

Speed is another issue on which industrial and academic partners do not see eye to eye. Academics take a long view, with typical projects lasting a few years — long enough to complete a PhD. Industrialists run a tighter ship, with stringent deadlines set by company financial-reporting timescales and market competition, often measured in months. Releasing a product onto the market a year after a competitor's is simply not an option.

"Business success for a product — the return on research investment — must take place within three years for a manager to profit from it," nanotechnologist Robert Brunner told me; he joined the University of Applied Sciences Jena in Germany in 2010 after many years in industry. "A proof of principle is not enough; a market-ready product must be there." Agreeing on timeline expectations is as essential as agreeing on technical specifications.

So, is there any use in all the science parks and technology 'hatcheries' with which universities have tried to attract companies onto campus? I think not, although evidence is equivocal. Science parks provide infrastructure for companies, especially for start-ups. But the celebrated 'proximity-to-campus' is usually only physical, with little real promotion of intimate collaboration between people.

For example, the Engineering Research Centers (ERCs) initiated by the US National Science Foundation in the 1980s to encourage cross-disciplinary research failed to live up to expectations, owing to insufficient participation by both industry professionals and academics. Almost 70% of engineers surveyed felt that the ERC objective had been poorly met or had no impact on industry⁷. Proof of Concept Centers⁴, the latest campus approach implemented in the United States, are intended to provide a spectrum of services to help to disseminate technology from university to industry. But the centres await a precise definition of their role.

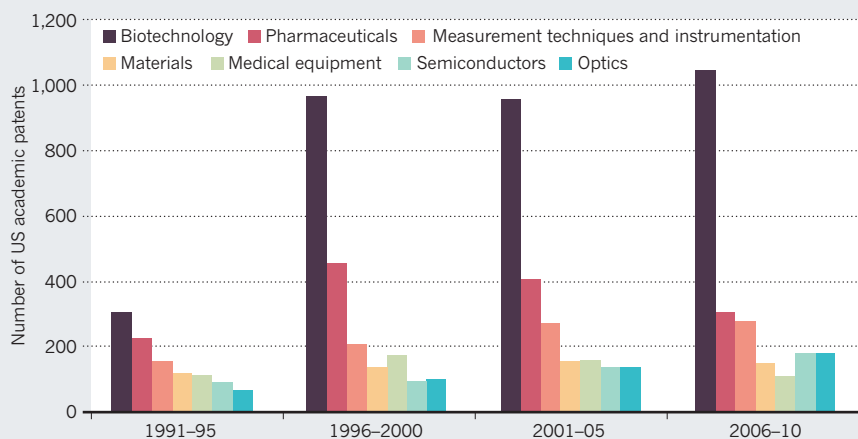
WORK TOGETHER

To stop promising technologies languishing on laboratory shelves, universities need to realize that industry is not a rapacious octopus, sucking up everything it can get its tentacles on and suffocating the scientific independence of academic researchers. Patents and industrial research should be valued more highly in faculty evaluations, and university liaison offices must be willing to

"Diverging technical expectations mar many industrial-academic relationships."

PATENT PROBLEM

Physical-sciences patents in the United States have increased, but biomedical inventions still dominate.



SOURCE: US NSF/THE PATENT BOARD

sell emerging technology and intellectual property at a reasonable cost to support product development. Academics must realize that just sending off a progress report after cogitating in the laboratory is not useful for a company.

On the other side, industry must appreciate that universities are not buffets of fully mature technologies there for the taking, free of charge. Companies should expect to invest time and money to move from an academic prototype to a commercial product. Industrialists must acknowledge that repeated measurements of reproducibility, lifetime and reliability are difficult to fit into the academic framework of constant innovation, and are not the best use of researchers' skills.

What both partners need from each other must be made clear at the outset. As Olav Solgaard, an electrical engineer at Stanford, explained: successful collaborations require leaders on both sides to manage expectations and to set sensible ground rules. Both must agree on outputs such as publications, especially when PhD students are involved. Simple and direct approaches are necessary. Many universities in China, for example, are involved in managing the companies with which they collaborate⁸. Others find that integrating industrial researchers into university laboratories is effective.

Hiroshi Toshiyoshi at the University of Tokyo, who has a long record of developing microelectromechanical systems and collaborating with industry told me how he likes to operate: "I like to ask my partner company to send their researchers to my group, where I give theoretical and on-the-job training for a year or two. We may not be able to deliver immediate results, but the company will obtain long-lasting competence."

So, set aside some lab space, fill it with recent graduates and company researchers, shake well and let the nutty academic idea evolve into the useful industrial prototype. ■

Hans Zappe is chair of micro-optics in the Department of Microsystems Engineering, University of Freiburg, Germany.
e-mail: zappe@imtek.uni-freiburg.de

1. Okamuro, H. & Nishimura, J. J. *Technol. Transf.* **38**, 273–301 (2013).
2. Grimpe, C. & Fier, H. J. *Technol. Transf.* **35**, 637–650 (2010).
3. Czarnitzki, D., Glänzel, W. & Hussinger, K. *Res. Eval.* **16**, 311–319 (2007).
4. Bradley, S. R., Hayter, C. S. & Link, A. N. *J. Technol. Transf.* **38**, 349–381 (2013).
5. Kenney, M. & Goe, W. R. *Res. Policy* **33**, 691–707 (2004).
6. Perkmann, M. *et al. Res. Policy* **42**, 423–442 (2013).
7. Marazita, C. F. *Technol. Forecast. Soc. Change* **39**, 397–410 (1991).
8. Wu, W. *World Dev.* **35**, 1075–1093 (2007).



An endangered Florida panther population was bolstered through hybridization with a related subspecies — a technique that could be refined using genomic tools.

Gene tweaking for conservation

It is time to weigh up the pros and cons of using genetic engineering to rescue species from extinction, say

Michael A. Thomas and colleagues.

Even the most conservative estimates predict¹ that 15–40% of living species will be effectively extinct by 2050 as a result of climate change, habitat loss and other consequences of human activities. In the face of such drastic losses, scientists are debating the pros and cons of various, and often controversial, interventions. These include moving populations to help track hospitable habitats, and reinstating keystone species — those that have a large effect on ecosystem structure and function, such as top-level predators — into areas where they have long been absent^{2,3}. Even the revival of species that have recently gone extinct is being explored.

So far, an increasingly viable (and potentially less risky) option, which we call facilitated adaptation, has been little discussed. It would involve rescuing a target population or species by endowing it with adaptive alleles, or gene variants, using genetic engineering.

Over the past 30 years, genetic engineering in agriculture has received substantial

attention. Today, 12% of arable land worldwide is planted with genetically modified (GM) crops; the GM seed market alone is valued at US\$15 billion. As techniques become ever more sophisticated, more possibilities will open up.

We believe that these combined factors mean that conservationists will almost certainly be tempted to apply genetic engineering to safeguard biodiversity. Facilitated adaptation might be less logistically challenging than moving entire populations, and less fraught with ecological and socio-economic complications — relocation could introduce harmful invasive species, for example, or unleash outbreaks of disease. But facilitated adaptation is likely to be beset with other challenges and pitfalls. Now is the time to consider what those might be.

THREE OPTIONS

There are at least three ways to avert extinction using facilitated adaptation. First, animals or plants from a threatened

► population could be crossed or hybridized with individuals of the same species from better-adapted populations to introduce beneficial alleles into the threatened population. Second, specific alleles drawn from a well-adapted population could be directly transferred into the genomes of threatened populations of the same species. And third, genes taken from a well-adapted species could be incorporated into the genomes of endangered individuals of a different species. Each approach carries its own set of challenges, complications and risks.

BACK FROM THE BRINK

Conservationists have already tried hybridizing individuals to aid the recovery of vulnerable populations. For instance, in an isolated population of the viper *Vipera berus* in Sweden, the number of inviable offspring produced as a result of inbreeding plummeted when male vipers from a healthy population were introduced⁴. Similarly, the size of a remnant Florida panther (*Puma concolor coryi*) population effectively increased by 100% after conservationists introduced eight cats of a related subspecies, *Puma concolor stanleyana*, from Texas. Heterozygosity, a measure of genetic variation, also rose in the Florida population, from around 18% to more than 25% in 12 years⁵.

This practice of injecting lost genetic diversity into a threatened population could be refined using genomic tools. For instance, conservationists could screen potential donor populations for individuals with alleles that would, say, yield physiological tolerance of warmer temperatures, or resistance to a catastrophic disease.

The potential risks of this approach, as with relocating entire populations, include the introduction of wildlife diseases, the dilution of locally adaptive alleles and the disruption of co-adapted gene complexes that impart advantages crucial in the threatened population's local habitat.

Success with the second approach — directly transferring specific alleles drawn from a well-adapted population into individuals from a threatened population — would similarly depend on introducing enough 'better-adapted' individuals with a sufficient selective advantage to increase the mean fitness of the threatened population. This would require conservationists first to work out which genes are controlling relevant adaptive traits. Earlier this year, aquaculture researchers identified⁶ alleles associated with heat tolerance in a commercial rainbow trout, *Oncorhynchus mykiss*. These gene variants could potentially be inserted into the genomes of fish eggs or embryos in populations threatened by rising water temperatures because of climate change.

This strategy has promise if one or a few

genes have a drastic impact on the trait of interest. For instance, some amphibians are resistant to the fungal disease chytridiomycosis, which has been linked to massive population declines and several extinctions. If one or a few genes are found to increase resistance, they might make excellent targets for transfer.

This form of genetic engineering is probably the least risky of the approaches described here, because it involves the

"Well-adapted combinations of alleles could be disrupted."

movement of alleles within the same species, without the attendant risks of moving diseases or parasites. But again, the challenge is that numerous genes tend to be associated with complex traits. Interactions between genes and the environment, and between introduced alleles and existing genetic diversity, will also be difficult to predict. And, as with the hybridization approach, well-adapted combinations of alleles could be disrupted.

The third transgenic approach — transferring genes between different species — has been used for more than a decade to improve resistance to crop stressors, including drought and extreme temperatures⁷. Certain genes from rice and the cress plant *Arabidopsis*, for instance, make tomato plants more tolerant of cold temperatures. A few years ago, developmental biologists produced mice with exceptionally long digits by replacing a limb-specific transcriptional enhancer of the mouse *Prx1* gene locus with the homologous bat sequence⁸. Conservationists could use similar approaches to explore the use of other genes, to augment, for example, resistance to white-nose syndrome, a rapidly spreading and deadly fungal disease in North American bats.

The effects of moving genetic variants between species, however, are likely to be even harder to predict than those of transferring variants within species, and a major concern is that such an approach could bring unintended and unmanageable consequences.

PRESERVATION PROGRAMMES

To begin the process of identifying adaptive genes and alleles, and predicting how they will behave in the target genomic and environmental contexts, conservationists can tap several resources. Natural history museum and herbarium collections worldwide, for example, can provide basic data on biogeography and genetics for many animal and plant species.

Large-scale preservation projects could provide the actual tissues needed to explore adaptations among organisms in different ecological contexts. One such project is the

Global Genome Biodiversity Network — an effort to preserve and provide open access to genomic information and DNA samples from various collections, including those of the Smithsonian National Museum of Natural History in Washington DC (see www.mnh.si.edu/ggi).

Ultimately, successful facilitated adaptation will require unprecedented collaboration between organismal, ecological and molecular biologists and climate scientists⁹. Biorepositories — such as seed banks, natural history museums and zoological parks, including the Frozen Zoo at San Diego Zoo in California, which houses around 9,000 frozen cell samples from endangered species — will need to be integrated with advances in biotechnology and efforts to explore the genomic mechanisms underlying adaptive traits associated with climate change¹⁰, catastrophic diseases and so on.

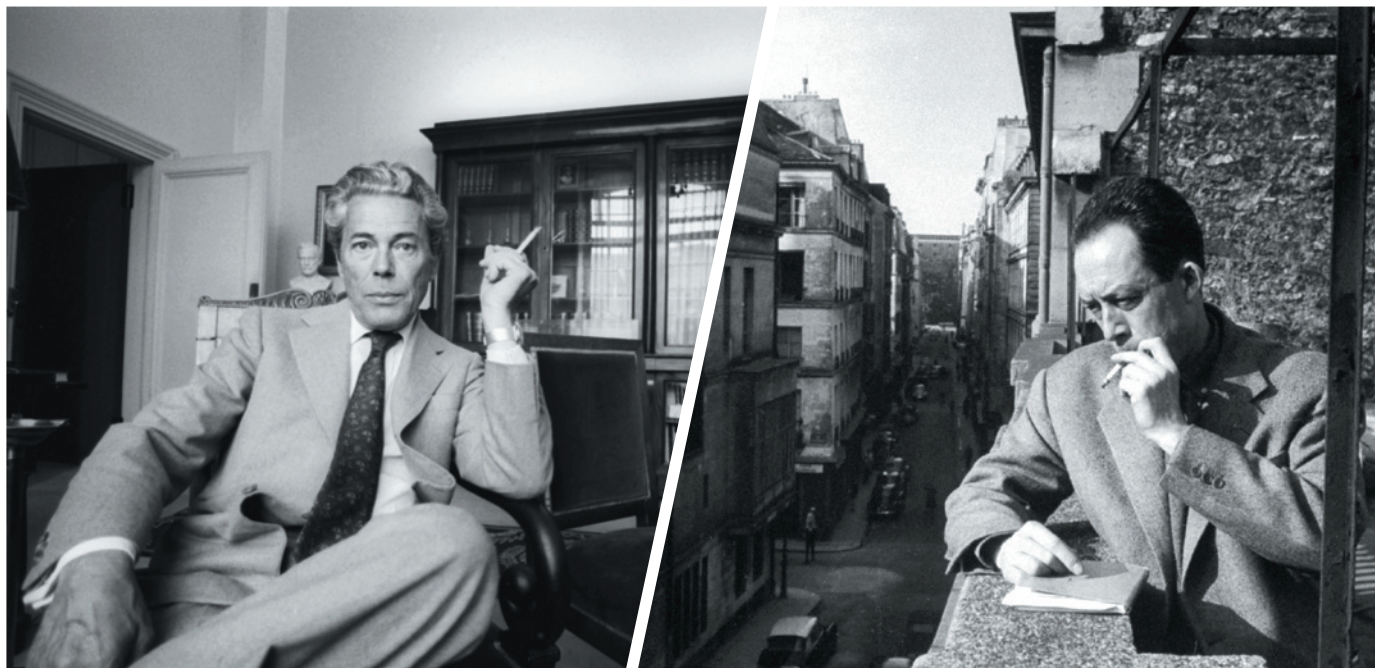
Facilitated adaptation will also require a change in people's views about biodiversity conservation and its ethics, practices and impact on society. Even moving members of the panther subspecies from Texas into the Florida panther's ecosystem incited misgivings over the appropriateness of meddling directly with organisms rather than with their habitat.

A serious concern is that even the possibility of using genetic-engineering tools to rescue biodiversity will encourage inaction with regard to climate change. Before genetic engineering can be seriously entertained as a tool for preserving biodiversity, conservationists need to agree on the types of scenario for which facilitated adaptation, managed relocation and other adaptation strategies might be appropriate, and where such strategies are likely to fail or introduce more serious problems.

For some species, facilitated adaptation could turn out to be the only viable remedy. ■

Michael A. Thomas is professor of biology at Idaho State University in Pocatello, Idaho.
Gary W. Roemer, C. Josh Donlan, Brett G. Dickson, Marjorie Matocq, Jason Malaney.
e-mail: mthomas@isu.edu

1. Thomas, C. D. & Williamson, M. *Nature* **482**, E4–E5 (2012).
2. Donlan, J. et al. *Nature* **436**, 913–914 (2005).
3. Hoegh-Guldberg, O. et al. *Science* **321**, 345–346 (2008).
4. Madsen, T., Stille, B. & Shine, R. *Biol. Conserv.* **75**, 113–118 (1996).
5. Johnson, W. E. et al. *Science* **329**, 1641–1645 (2010).
6. Rebl, A. et al. *Mar. Biotechnol.* **15**, 445–460 (2013).
7. Zhang, J. et al. *In Vitro Cell. Dev. Biol. Plant* **36**, 108–114 (2000).
8. Cretekos, C. J. et al. *Genes Dev.* **22**, 141–151 (2008).
9. Redford, K. H., Adams, W. & Mace, G. M. *PLoS Biol.* **11**, e1001530 (2013).
10. Whitehead, A. J. *Exp. Biol.* **215**, 884–891 (2012).



Both French Resistance firebrands in the war-torn 1940s, molecular geneticist Jacques Monod (left) and writer Albert Camus later became friends.

MOLECULAR GENETICS

A revolutionary meeting of minds

Jan Witkowski relishes the interwoven stories of Nobel laureates Jacques Monod and Albert Camus.

In *Brave Genius*, the biologist Sean Carroll tells the stories of two remarkable men whose lives intersected in the aftermath of the Second World War, sparking a rare communion of spirit. Jacques Monod was a founder of molecular genetics who, with biologists André Lwoff and François Jacob, received the 1965 Nobel Prize in Physiology or Medicine. Novelist and philosopher Albert Camus was another Nobel laureate, receiving the 1957 literature prize for work “which with clear-sighted earnestness illuminates the problems of the human conscience in our times”. Just over two years later, Camus died in a car crash.

Both men were fearless in their fight against the German invaders; both were geniuses in their own spheres; and the two were united in their world views. Monod’s scientific outlook informed Camus’s work, and Camus’s philosophy greatly influenced Monod, so much so that Carroll calls Monod “Camus in a lab coat”.

Carroll’s story takes off in Paris in 1940, the year the city fell to the Germans. Monod

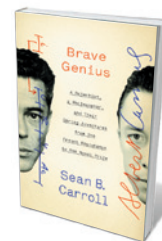
was part-way through his thesis at the Sorbonne, studying the growth of bacteria. That year he discovered ‘adaptation’, the lag that occurs when bacteria that are growing in a medium with two sugars exhaust the first and switch to the second. The observation led to his life’s work exploring the regulation of gene expression in bacteria. Around the same time, Camus arrived in Paris from Algeria. Already an experienced reporter and editor, he found a job at the newspaper *Paris-Soir*, but he had grander ambitions: to write about the absurd. In this philosophy, the Universe has no goal or meaning. Far from being a credo of despair, it urges people to recognize the absurd and live life to the full.

As Carroll shows, each man heeded French general Charles de Gaulle’s call that “all free Frenchmen, wherever they be, should continue the fight as best they may”. Monod distributed *Résistance*, a clandestine newspaper, and later joined a group that sabotaged German supply lines. He played a leading part in the rebellion of

August 1944, when the Resistance fought German troops in the streets of Paris. Camus joined a Resistance cell and wrote editorials for its newspaper, *Combat*. Carroll brings out how each man managed to pursue his vocation under the most dangerous of circumstances. Monod moved to a now legendary attic laboratory at the Pasteur Institute with Lwoff. Meanwhile, Camus continued to write, publishing *L’Étranger* (*The Outsider*) in 1942 and establishing his place among the French intelligentsia with his then friends Jean-Paul Sartre and Simone de Beauvoir.

It was a shared revulsion at Soviet biologist Trofim Lysenko’s genetics theory that brought Monod and Camus together. In August 1948, the Paris Communist paper *Les Lettres Françaises* extolled Lysenko’s work as “A Great Scientific Event”. Three weeks later, Monod wrote a counterblast

in *Combat*, declaring Lysenko’s work non-scientific. By this time, Camus had come to believe that any totalitarian regime, including those of communist states, was inherently corrupt — a view confirmed by conversations with the Hungarian-born writer Arthur Koestler, whose anti-totalitarian novel *Darkness at Noon* had infuriated communists. Camus was working on *L’Homme* ▶



Brave Genius: A Scientist, a Philosopher, and Their Daring Adventures from the French Resistance to the Nobel Prize
SEAN B. CARROLL
Crown: 2013.

► *Révolté (The Rebel)*, his long essay on the nature of revolution. He was introduced to Monod and, as Carroll puts it, they “hit it off right away”. Monod helped Camus with a chapter on how communism had perverted science in the Soviet Union, exemplified by Lysenko. (The cruel Soviet suppression of Hungary’s 1956 uprising put the seal on their mutual rejection of communism.)

Camus’s influence on Monod is clear in *Le Hasard et La Nécessité (Chance and Necessity; Seuil, 1970)*, the international best-seller in which Monod drew on the new science of molecular biology. Biology, for instance, had demonstrated that *Homo sapiens* arose through a series of chance events and that there is no grand design to the Universe. In a reference to Camus’s world view, Monod wrote that man “lives on the boundary of an alien world; a world that is deaf to his music, and as indifferent to his hopes as it is to his suffering or his crimes”. In the late 1950s, Monod carried out experiments with François Jacob and biochemist Arthur Pardee that hinted at the existence of an unstable intermediate between DNA and ribosomes. Further work by Monod and Jacob led to the operon model of how gene expression is regulated, described in a classic paper in 1961. The same year Jacob, with biologists Sydney Brenner and Matt Meselson, provided the experimental confirmation by demonstrating the existence of messenger RNA. The Nobel prize for Monod, Jacob and Lwoff followed.

The journalist Jean Daniel observed of the comradeship between Monod and Camus that there was “a complicity so intense ... that only a shared kindness of heart allowed them not to find unwelcome those who interfered in their privacy”. I am not sure that Carroll has conveyed that intensity — perhaps no one can. But although *Brave Genius* is a long and complex book, Carroll does a masterful job of keeping the many elements together and the story moving. I learned much about France at the time of the Second World War, and was prompted to reread Camus’s great novel *La Peste (The Plague)*.

In 1959, C. P. Snow wrote of the “two cultures” — that gulf between science and the humanities. *Brave Genius* provides an opportunity for those on both sides of the divide to sample a potent mix of genetics, philosophy and literature, forged in the twentieth-century tumult of war and cold war. ■

Jan Witkowski is executive director of the Banbury Center, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA.
e-mail: witkowski@cshl.edu



Winston Churchill on a voyage across the Atlantic in October 1941.

ATOMIC SCIENCE

Winston and the warheads

Richard Rhodes explores a history of Britain’s little-known role in the race to develop an atomic bomb.

Britain and the United States, which cooperated so effectively as military allies during the Second World War, collaborated only intermittently — and from the British point of view inadequately — in the development of the first atomic bombs. The US side of the story has been told more than once; the British side, not recently explored, is now tackled by Graham Farmelo in *Churchill’s Bomb*.

The author, a physicist, ranges across Winston Churchill’s long career — from 1901, when Churchill wrote to H. G. Wells to congratulate him on *Anticipations*, a work of predictive non-fiction, to his final turn as prime minister in the early 1950s, when he pushed for a British hydrogen bomb. Farmelo is especially good on the Second World War years, revealing much about the Anglo-American relationship that has been guarded or unclear.

British work on the bomb preceded that by the United States. Britain was at war for more than two years before the United States came in, and was inevitably more urgently concerned with German uranium research. Moreover, Britain’s generous policy of taking in refugee Jewish scientists who were



Churchill’s Bomb:
A Hidden History of Science, War and Politics
GRAHAM FARMELO
Faber & Faber: 2013.

fleeing the Nazis supplied a cadre of highly motivated physicists to investigate the explosive properties of uranium at a time when most British physicists were working on radar. In fact, it was the refugee scientists who first alerted the British government to German uranium research, just as their US counterparts

famously enlisted Albert Einstein to alert President Franklin Roosevelt.

On both sides of the Atlantic, however, gatekeeper scientific advisers delayed progress. In the US case, the culprit was a government scientist named Lyman Briggs. Briggs, the director of the National Bureau of Standards, so overemphasized secrecy that the meeting minutes he received from the MAUD committee — the group of British officials tasked with researching the feasibility of building an atomic bomb — languished

EVERETT COLLECTION HISTORICAL/ALAMY

undistributed in his safe. Eventually, Briggs was moved aside and his safe opened to reveal its treasures.

In Britain, Farmelo reports, the problem was more serious because the gatekeeper was Winston Churchill's personal scientific adviser, Frederick Lindemann. Lindemann did not quite believe in the bomb, and in any case thought it should be built in Britain or, if that proved impossible, Canada.

At a crucial point in Churchill's ongoing negotiations with Roosevelt, in October 1941, a message arrived from the US president offering to coordinate "or even jointly conduct" a bomb programme. British official opinion still favoured consultation between parallel projects, Farmelo notes, rather than full collaboration. Discouraged by Lindemann and other advisers, Churchill delayed his response. Almost two months passed before he answered the president's note, and even then he did so only tepidly.

Those two months were crucial: the US programme officially expanded to full industrial scale on 6 December 1941, and the following day the Japanese bombed Pearl Harbor, shocking the United States into joining the war. The atomic-bomb programme was soon assigned to the US Army Corps of Engineers and, a few months later, now renamed the Manhattan Project, put under the command of a big, dynamic, no-nonsense engineering general named Leslie Richard Groves, a combative Anglophobe.

Curiously, Groves and his government superior, the science czar Vannevar Bush, worried as much about giving Britain a leg up on post-war nuclear-power development as they did about sharing the 'secrets' of the bomb. Early in 1943, Farmelo writes, Groves blocked British participation almost entirely after Imperial Chemicals' Wallace Akers was chosen to run the British programme. A secret Anglo-Russian agreement to share new and future weapons further soured Groves and Bush when they learned of it, although Farmelo argues that Churchill would have repudiated the agreement instantly had he thought it would quash collaborations with the United States. By the time all these misunderstandings had been sorted out, the British lead was buried in the dust of Groves' multi-pronged, multibillion-dollar race to an almost all-American bomb.

Farmelo's book sometimes falters on technical details. Breeding plutonium in large uranium-graphite reactors in eastern Washington state becomes "the production of chemicals containing weapons-grade plutonium". The bomb tested in the New Mexican desert in July 1945 was not, as he writes, "the first nuclear bomb" — that was the uranium gun bomb, Little Boy, already sailing towards Japan — but rather the first plutonium implosion assembly, the 'Gadget', its technology so radically new it needed ►

Books in brief



Command and Control: Nuclear Weapons, the Damascus Accident, and the Illusion of Safety

Eric Schlosser PENGUIN (2013)

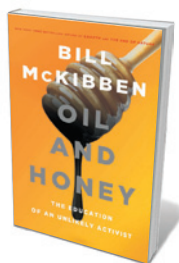
After dishing the dirt on junk food in *Fast Food Nation* (Houghton Mifflin Harcourt, 2001), Eric Schlosser tackles another weapon of mass destruction: the US nuclear arsenal. His propulsive narrative alternates between a history of nuclear arms and an account of a near miss: the explosion of a Titan II ballistic missile in Arkansas. Invoking sociologist Charles Perrow's finding that such "tightly coupled", interactive systems can be undone by trivialities, Schlosser calls for new thinking on this legacy — and soon.



Neurocomic

Matteo Farinella and Hana Roš NOBROW (2013)

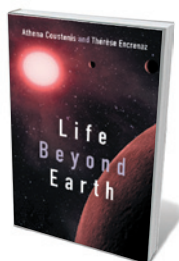
Illustrator Matteo Farinella and neuroscientist Hana Roš have crafted a graphic introduction to the human brain that seethes with many-layered invention. Boy meets girl, and is propelled into the *Alice in Wonderland* world of her brain — where, for example, Camillo Golgi and Santiago Ramón y Cajal grapple in a forest of neurons. Morphology and plasticity, for example, are distinct regions of 'Brainland', in which greats in each field are tour guides, and neural phenomena appear as anything from key-wielding superheroes (neurotransmitters) to a haunted castle (consciousness).



Oil and Honey: The Education of an Unlikely Activist

Bill McKibben TIMES BOOKS (2013)

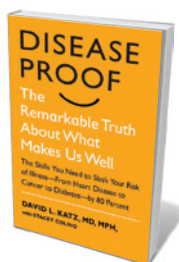
Environmental writer Bill McKibben set off shock waves in 1989 with *The End of Nature* (Anchor), the first popular treatment of climate change. Two years ago, with the impacts of oil-industry 'business as usual' biting, McKibben moved on to full-time climate activism. In this eloquent memoir, he interweaves reportage on deluges, heatwaves and melts with demonstrated solutions to "malfunctioning modernity". High-profile protest is only part of that, he argues. A revolution in local sustainability is also essential — and achievable, as the story of a Vermont bee-keeper reveals.



Life Beyond Earth: The Search for Habitable Worlds in the Universe

Athena Coustenis and Thérèse Encrenaz CAMBRIDGE UNIVERSITY PRESS (2013)

Is the biosphere that so astounds us one of thousands? In this packed primer on exoplanetary life, distinguished astrophysicists Athena Coustenis and Thérèse Encrenaz summarize the science and speculation. Kicking off with planet formation, life's origins on Earth and extreme environments, they boldly go into areas such as potential habitats in the outer Solar System and far-future ideas such as "terraforming" Mars for human habitation.



Disease-Proof: The Remarkable Truth About What Makes Us Well

David L. Katz and Stacey Colino HUDSON STREET PRESS (2013)

We are living longer, chronic disease rates are rising — and the trade-off is a dubious one, argues medic David Katz. With writer Stacey Colino, Katz presents a research-based approach to increasing the chances of reaching a healthy old age that brims with intelligent suggestions for behavioural change. The discipline involved is workable, from eliminating 'stealth' sources of sugar to exercising in the 'nooks and crannies' of a busy day; and Katz's vision for wellness encompasses societal change. *Barbara Kiser*

► to be tested at full yield. Nor did the scientists assembled in the desert to watch that first test apply suntan lotion to protect themselves from “the radiation blast”. It was the high-intensity light from the nuclear fireball that concerned them.

More egregiously, Farmelo misses what is in my view a crucial part of the post-war negotiations between the United States and Britain over uranium supplies. The United States was at that time believed, for reasons I have never understood, to have only modest domestic sources of uranium ore. The two countries had agreed during wartime that

they would share the rich ore resources of the Belgian Congo equally. By late 1947, Britain was approaching bankruptcy, a congressional debate neared on the Marshall Plan and several conservative US senators had been outraged to learn that Britain still had a veto over any US use of atomic bombs. The administration of President Harry Truman demanded changes: Britain would give up its veto as well as its share of the Belgian Congo ore; the United States, in return, would continue to aid its wartime ally economically. It was this ore grab — formalized in a *modus vivendi* of 7 January 1948 — not bureaucratic

dithering, that delayed the British bomb.

Churchill's Bomb is colourful but incomplete, focused more on Churchill than on the bomb. It is a useful adjunct to what is still the best series on the British bomb, Margaret Gowing's official history *Britain and Atomic Energy 1939–1945* (Macmillan, 1964) and its successor volumes. ■

Richard Rhodes is the author of the Pulitzer-prizewinning *The Making of the Atomic Bomb* and three further volumes of nuclear history.
e-mail: rhodes.today@comcast.net

MILITARY SCIENCE

Fight in flight

Ann Finkbeiner ponders a chronicle of airborne war.

To see farther, go higher: from horseback, hilltop and tower in the eighteenth century to balloons in the nineteenth, aeroplanes and satellites in the twentieth and robotic drones in the twenty-first. With each step up in height and technology comes a broader view of enemy territory and a greater personal distance from it. What to make of this?

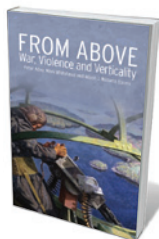
In *From Above*, the view from higher up translates into a greater power to acquire and rule, to control and to kill. The 13 authors in this collection of essays, edited by Peter Adey, Mark Whitehead and Alison Williams, are academic humanists and social scientists linked by an interest in how human interaction with geography has shaped warfare.

The essays, which are divided into three categories, are often built around case studies and begin with the view from the sky. As captured in drawings, photographs and film for much of the past century, images taken from above tend to be visually confusing and must be interpreted and even manipulated. The imaging of large swathes of territory requires the formation of photomosaics, in which photographs taken at different times, heights and angles are stitched together. The slight sense of unreality inherent in the view from above, many of the authors argue, contributes to an emotional distance; the result is that conquering or killing becomes easier.

Meanwhile, warfare itself has changed: war is now fought not by vast, easily bombable armies but by small groups of insurgents who are hard to spot from the air. In the Vietnam War, for instance, seismic and acoustic sensors on the ground were used to locate insurgents. When triggered, the sensors

signalled to distant computers that calculated, then sent, the enemy's coordinates to high-flying bombers. That war, writes geographer Derek Gregory (quoting from an article by Paul Dickson and John Rothchild), was a “lethal pinball machine” that — fast-forward to the wars in Iraq and Afghanistan — became a network of surveillance and targeting drones run by people who commute to work. The more distant the killing, the more impersonal, and the more the exercise resembles a video game.

The second group of essays focuses on the responses of those on the ground to being viewed from above — the immediate reaction being concealment. In the Second World War, for example, when the United Kingdom wanted to camouflage its industry and infrastructure from German bombers, people learned to think of the landscape as seen from the air. The Home Office even ran an unsuccessful experiment in which oil tanks were disguised with green and brown paints of differing reflectivity to harmonize with the British landscape. Another on-the-ground reaction is to spy on the sky. One international group of people practises a “peculiar version of amateur astronomy”. Using little more than good binoculars, stopwatches, star charts and Kepler's laws of planetary motion, they track highly classified reconnaissance satellites. The



From Above: War, Violence and Verticality
EDITED BY PETER ADEY, MARK WHITEHEAD AND ALISON J. WILLIAMS
Hurst: 2013.

satellites are usually reflective, so although no government admits to their existence, they are trackable. Their orbits reveal where they are going and a little of what they are doing — including when they fall out of the sky.

The third set of essays covers interaction between the sky and ground. Bombers are frightening because of their purpose, so their very presence in the sky is intimidating: one aim of bombing runs has always been to undermine morale. The 2003 campaign in the US war with Iraq was explicitly called Shock and Awe because it aimed to sap the Iraqi will to fight. The interaction between air and ground is most easily seen in the use of unmanned surveillance drones. Each drone needs four people to guide it and to keep track of its technologies and communications — which they do from many miles away. Drones return vast amounts of information. If aerial views began with a person climbing a hill and then climbing back down to analyse what was seen, then drones almost seem to conflate person, view and action.

From Above is written by academics for academics. The case studies are fascinating, but the sentences are often opaque. (In one example, an author discusses the ‘weaponization’ of the cinema, writing that it has “particular capacities for movement whose influences on specific ideas of global escalation make them into logistics of perception or the escalation of the modern technical beyond”.) Thus, the ideas and connections between them are frustratingly hard to understand.

I think the book's main message is that the aerial view confers a remoteness that enables violence. Implicit throughout, as stated in the introduction, is the judgement that in spite of the “spectacle and beauty” of the aerial view, “we must be careful not to celebrate it”. Since the first stone tools, technologies have had dual uses, both civilizing and military, and we should remember that duality. ■

Ann Finkbeiner is a freelance writer in Baltimore, Maryland, and author of *The Jasons*.
anniefk@gmail.com

TECHNOLOGY

Pulp, pilcrows and interrobangs

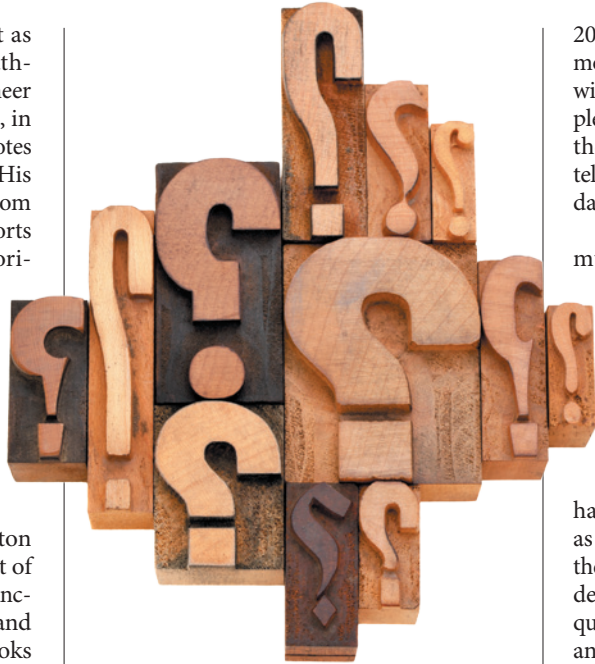
Andrew Robinson savours a pair of lively studies on paper and punctuation.

“The paperless society is about as plausible as the paperless bathroom”, wrote Jesse Shera, a pioneer of information technology in libraries, in 1982. Nicholas Basbanes approvingly quotes this remark near the end of *On Paper*. His edifying, if bloated, history of paper — from manuscripts, books, newspapers, passports and currency notes to stationery and origami, packaging, cigarettes and toilet paper — has also convinced me that, despite digitization, Shera was right. The stuff is convenient, portable and cheap: *On Paper* has been published, after all, as both an e-book and a deckle-edged hardback.

Typography, of course, enables most paper-based and digital communication. In *Shady Characters*, Keith Houston celebrates the origins and development of typographical symbols, particularly the punctuation marks that give shape, rhythm and sense to the written sentence. These books offer much insight into the centuries of invention that have gone into creating the norms with which we unthinkingly consume and communicate vast amounts of information.

Basbanes begins with a visit to the paper makers of remote, mountainous, southwestern China, whose ranks are rapidly thinning. Tradition has it that a Chinese court eunuch, Cai Lun, invented papermaking in AD 105, but the process — pulping the cellulose fibres of rags — was probably pioneered several centuries earlier. (The Egyptian manufacture of papyrus long predates paper, but involves lamination rather than pulping. Confusingly, the word ‘paper’ is derived from the Latin *papyrus*.) The technology then travelled eastwards to Korea and Japan, and westwards along the Silk Road through Central Asia, via the Arabs, to Europe. In both directions, Buddhist monks were the first to use the material to record sacred texts. The world’s earliest complete survival of a dated printed book is the Buddhist *Diamond Sutra*, published in 868, discovered a century ago in a cave at Dunhuang, western China, on the Silk Road. It is now in the British Library in London.

One country after another adopted paper, from Spain in 1056 and Germany in 1391 (not long before Johannes Gutenberg began printing) to North America in 1690 and Australia in 1818. From the mid-nineteenth century, paper made from wood pulp became a reality and, in 1873, *The New York Times* led the conversion of nearly every US newspaper from rag paper to wood-pulp newspaper.



On Paper: The Everything of Its Two-Thousand-Year History

NICHOLAS A. BASBANES
Knopf: 2013.

Shady Characters: The Secret Life of Punctuation, Symbols and Other Typographical Marks

KEITH HOUSTON
W. W. Norton: 2013.

With the rise of mass literacy at the end of the Victorian era, “pulp fiction” soon followed.

As Basbanes discusses, paper also became physically useful in wars. Gun cartridges were made of paper from the fourteenth century; the word ‘cartridge’ is probably derived from *cartouche*, which means ‘roll of paper’ in French. In the Second World War, the Japanese even made paper-balloon bombs of about 10 metres in diameter, each constructed from 600 sheets of handmade mulberry paper (*kozo*) glued together and filled with hydrogen gas. From late 1944 until April 1945, some 9,000 were launched across the Pacific on the jet streams. An estimated 1,000 reached the United States, but the sole casualties were a woman and five children in Oregon.

The intimate relationship between paper and script, and its dominance as a mode of communication — at least until the past 15 years or so — underpin *On Paper*. Basbanes memorably describes, for instance, the blizzard of office paper that emanated from New York’s Twin Towers during the 11 September

2001 attacks. A blood-stained sheet of common bond found at ground level was scrawled with the words: “84th floor west office 12 people trapped”. Ten years later, DNA testing of the blood identified the writer. His widow tells Basbanes that this sheet “belongs to my daughters. It’s their legacy from their father.”

Enlivening and amplifying our communications are the signs and symbols that typography blogger Houston unravels in his first book, *Shady Characters*. The e-mail staple @ merits an entire chapter; Houston also examines the dagger (†), the ampersand (&), the hyphen and quotation marks, as well as more obscure signs such as the pilcrow (¶) and the pictographic ‘manicule’ (a small hand with a pointing finger, once common as a textual highlighter but later relegated to the likes of Monty Python sketches). He even delves into the ‘interrobang’, a fusion of the question and exclamation marks invented by an advertising executive, Martin Specker, in the 1960s — which never really caught on.

Houston’s title refers to the mystery of how these “typographic conundrums” — as he calls them — have been endowed with so much meaning. Consider the hash sign, now much used in Twitter ‘hashtags’. It can denote a number (#5), weight (5#) or check-mate in chess; indicate a place to insert a space in proofreading; stand in for the sharp symbol in musical notation; or indicate, in many computer-programming languages, that the rest of the line is a comment and not a part of the program. The sign probably arose from the English abbreviation of the Latin word *libra*, or ‘scales’, as ‘lb’ (meaning a pound in weight). Scribes initially wrote ‘lb’ with a horizontal stroke through the two ascenders to indicate that it was a contraction; hasty writing eventually transformed that into the hash sign. An example of such a scribble by Isaac Newton is one of the book’s many intriguing illustrations.

Houston brings considerable wit and occasional erudition to the 5,000-year-old enigma of how we attempt to communicate our thoughts through visible signs. Like Basbanes’ doughty history, *Shady Characters* might make you look at books, or even this journal — in print or online — in an entirely new way. ■

Andrew Robinson is author of *The Story of Writing: Alphabets, Hieroglyphs and Pictograms*.
andrew.robinson33@virgin.net

MAREK ULJASZ/ALAMY

Correspondence

Citations: in defence of Brazilian journals

As editors of the Brazilian journal *Reports in Public Health* and signatories to the San Francisco Declaration on Research Assessment, we wish to point out that there are only four Brazilian journals among the total of 66 banned from Thomson Reuters' Journal Citation Reports (see go.nature.com/ufgbrn and *Nature* **500**, 510–511; 2013).

According to Brazil's open-access portal SciELO, there are more than 300 high-quality Brazilian scientific journals; these are also indexed in the main global bibliographical databases in their respective areas. They each contribute to the development of science worldwide.

This context should have been emphasized in your report, to avoid tainting the reputation of these many respected Brazilian journals by implication.

Marília Sá Carvalho, Claudia Travassos, Claudia Medina Coeli Reports in Public Health, Rio de Janeiro, Brazil.
cadernos@ensp.fiocruz.br

Citations: ethical ways to grow impact

As editors of a journal with one of the highest impact factors in Brazil, *Natureza & Conservação* (the Brazilian Journal of Nature Conservation), we suggest that there are more imaginative, ethical and effective ways to increase these ratings than the cross-citation methods used by a cluster of discredited Brazilian journals (see *Nature* **500**, 510–511; 2013).

In the past three years, our journal's impact factor has risen from 0.27 to 1.47. This is a result of inviting submissions from respected scientists worldwide to improve our visibility. We also recruit editors who have served on reputable international journals

to guarantee the quality of accepted papers.

Brazilian science is steadily improving in quality and quantity, thanks in part to strong support from two national funding agencies, the National Council for Scientific and Technological Development and the Coordination for the Improvement of Higher Education Personnel (R. D. Loyola *et al. Trends Ecol. Evol.* **27**, 585; 2012).

These agencies have set up committees of researchers from different fields to determine how impact factors can be used effectively to evaluate graduate programmes and to rank journals. In the process, they have standardized mean impact factors to take into account impact-factor variations between subdisciplines, thereby rendering comparisons between journals more meaningful.

Brazilian journals are ripe for development — all it takes is a rational approach to optimize their impact.

Rafael Dias Loyola, José Alexandre Felizola Diniz Filho Universidade Federal de Goiás, Goiânia, Goiás, Brazil.
loyola@ufg.br

Citations: overcome the language barrier

The 'citation stacking' practised by four Brazilian research journals (*Nature* **500**, 510–511; 2013) may be a misguided attempt to compensate for the recognition problems faced by journals that do not publish in English. It is in the interest of international science for researchers and funding agencies, as well as journals, to overcome such linguistic barriers.

Impact factors depend on the size of citation pools for different languages. This favours Chinese journals published in both Chinese and English, for example, but not Brazilian

journals that use Portuguese.

Transparent attempts to counteract this undue influence of language are under way. English-speaking scientists bidding for Chinese research grants, for example, enlist help from Chinese colleagues, with notable success. European nations such as Iceland and the Czech Republic require international and multilingual peer review of their research papers and grant applications — an approach that could also work for Brazilian research journals.

Ralf Buckley, Fernanda de Vasconcellos Pegas Griffith University, Gold Coast, Queensland, Australia.
r.buckley@griffith.edu.au
Zhong Lin-sheng Chinese Academy of Sciences, Beijing, China.

Food crisis spurs aid for poverty

Spikes in food prices in the past few years have heightened public awareness of global hunger and poverty, and have vastly increased funding for policies on agricultural development and food security. In short, we believe that the poor paradoxically stand to benefit from current high and volatile global food prices.

Mass media attention drives policy agendas (J. F. M. Swinnen and N. Francken *World Econ.* **29**, 637–654; 2006). The jump in food prices in 2008 attracted news coverage of hunger and farmers' problems in developing countries; such coverage has since kept pace with the volatile spiking in food prices.

This has resulted in more development aid from international organizations for agriculture and hunger prevention, reversing a long downward trend (U. Lele *Science* **327**, 1554; 2010). The private sector is also investing more in agriculture in developing countries.

We should approach today's food crisis as a valuable opportunity to combat global poverty and hunger.

Mara P. Squicciarini, Andrea Guariso, Johan Swinnen LICOS Centre for Institutions and Economic Performance, KU Leuven, Belgium.
mara.squicciarini@kuleuven.be

Once a scientist, always a scientist

Many scientists, particularly chemists, have careers away from a lab bench or professorial chair — but they are still scientists (*Nature* **500**, 369; 2013). As a mentor on careers for chemistry graduates (see go.nature.com/ajx23f), I consider myself a chemist for life.

Our training and experience mean that we are equipped to deal with the scientific issues that frequently crop up in alternative careers. This is true for scientists who have become teachers, information specialists, patent experts and practitioners, writers and consultants, to name but a few such alternatives. We continue to understand scientific terminology, to be adept at problem-solving and to think critically about scientific matters.

This could be one reason why the American Chemical Society in Washington DC recommends alternative careers in addition to more traditional paths (see go.nature.com/vs75vj).

Robert E. Buntrock Buntrock Associates, Orono, Maine, USA.
buntrock16@roadrunner.com

CONTRIBUTIONS

Correspondence may be submitted to correspondence@nature.com after consulting the author guidelines at <http://go.nature.com/cmchno>. Alternatively, readers may comment online: www.nature.com/nature.

BIODIVERSITY

Temperate hotspots

The inclusion of abundance data in global surveys of reef fishes reveals new hotspots of functional biodiversity, not all of which show high species richness. The findings may influence conservation priorities. [SEE LETTER P.539](#)



RICK STUART-SMITH

Figure 1 | Sea surveys. A volunteer Reef Life Survey diver moves along a transect to count reef fishes.

DEREK P. TITTENSOR

Our understanding of global patterns of biodiversity in the oceans has been founded primarily on the number of species in different regions. Although species richness is an important measure, a broader suite of metrics is desirable to fully integrate distinct aspects of ecological information and to detect locations that may not be species rich but are otherwise ecologically exceptional. Of particular interest is the diversity of functions among individuals and species, because numerous studies have linked this property to the goods and services that ecosystems provide, and to their resilience. On page 539 of this issue, Stuart-Smith *et al.*¹ present an impressive compilation of standardized diver-conducted underwater surveys that reveals intriguing patterns in the functional diversity of reef fishes around the world.

As any diver who has floated through the iridescent clouds of fishes that blanket tropical coral reefs will tell you, identifying them while underwater is a tricky proposition. Extrapolate this to also attempting to accurately assess their

abundance, and you have an idea of the effort that goes into underwater visual censuses (UVCs). UVCs are typically conducted either as stationary point counts, in which the diver stays in one place and identifies fishes within a circular area for a specified period of time, or as transects of specified length and width along which the diver swims.

There are then two possible approaches to compiling a worldwide data set based on UVCs. The first is to integrate existing data from the large number of such surveys that have already been conducted (see, for example, ref. 2). The advantage of using a large repository of surveys is obvious, but which statistical methods should be used to correct for differences in survey methodology is not. The other approach is to design a global standardized sampling programme from scratch, such as the Reef Life Survey used by Stuart-Smith and colleagues. This survey programme engaged researchers and 'citizen scientists' — volunteer scuba-divers who were trained and supervised by marine biologists — to generate more than 4,000 transect surveys at almost 2,000 sites around the world (Fig. 1).

The authors first used these data to analyse the species density (the relative richness per transect) and functional-group richness (the number of ecologically distinct groups) of reef fishes worldwide. Both of these variables show patterns that conform to typical latitudinal gradients — they decrease with increasing distance from the Equator. So far, so familiar. But surprises arose when they used the full power of their abundance data to assess species evenness (the distribution of individuals among species) and functional diversity (which integrates function and relative abundance). The authors found that species evenness increases towards the poles, running counter to the richness patterns. And the functional-diversity data display a curiously idiosyncratic pattern, peaking in areas that are not necessarily coincident with richness hotspots. In fact, some of these hotspots of functional diversity are out of the tropics entirely, with many being located in temperate boundary currents that flow towards the Equator, such as the Benguela Current. These currents are often characterized by high biological productivity driven by the upwelling of nutrient-rich water.

This study contrasts with previous analyses of functional diversity at global scales (for example, ref. 3), which have typically been terrestrial and founded on richness-based rather than abundance-based measures. The findings unlock a new view of the ocean in which high species richness does not necessarily correspond to elevated functional diversity. This concept complements a previous regional terrestrial study⁴ that also found a pattern of functional diversity that differed significantly from the pattern for species richness.

Stuart-Smith *et al.* propose two general categories for their functional-diversity hotspots. The first type is mostly located in temperate regions, has high species evenness and, although not necessarily species rich, shows a greater than average functional diversity for a given number of species. These hotspots could be considered the 'socialist communities' of the fish world — relatively equitable, with species fulfilling markedly distinct roles. The second category comprises tropical regions with high functional diversity and moderate species evenness; these are 'lands of opportunity', less equitable but with many more roles available.

The broader scientific implications of these findings will depend on how well reef fishes proxy for other taxa in terms of their spatial gradients. Patterns of species richness in coastal taxa are remarkably congruent⁵, but those in oceanic taxa are distinctly different from their coastal brethren. It will be informative for researchers to explore whether Stuart-Smith and colleagues' unexpected discoveries on evenness and functional diversity apply to other marine organisms. We also currently lack a testable mechanistic framework that can help to provide explanations for the processes that create the patterns of evenness and richness underlying the authors' observations.

Stuart-Smith *et al.* only touch on the relevance of these results for conservation and management, but this is clearly worthy of considerable discussion. Biologists and resource managers are placing increasing emphasis on ecosystem function, and overlaying the authors' map of functional diversity on that of marine protected areas around the world would be a straightforward first step towards assessing protection for function rather than solely for species richness. The study results also raise questions about what should be prioritized: centres of richness, which may also be engines of species generation⁶, or areas that are functionally fragile or functionally diverse, or that contribute disproportionately to human well-being.

The best of all possible worlds would be to safeguard all of the above. But our world is rather distant from such a Panglossian ideal, so this is likely to be too much to ask. Information such as that provided by Stuart-Smith *et al.* might not make our decisions any simpler, but the provision of comprehensive data on all aspects of marine biodiversity will at

least enable those decisions to be made on a stronger scientific footing. ■

Derek P. Tittensor is at *Dalhousie University, Halifax, Nova Scotia, Canada, the Microsoft Research Computational Science Laboratory, Cambridge, and the United Nations Environment Programme World Conservation Monitoring Centre, Cambridge CB3 0DL, UK. e-mail: derekt@mathstat.dal.ca*

ELECTRONICS

The carbon-nanotube computer has arrived

The most complex electronic device yet built from carbon nanotubes has been demonstrated. The system is a functional universal computer, and represents a significant advance in the field of emerging electronic materials. SEE LETTER P.526

FRANZ KREUPL

Transistors made from semiconducting carbon nanotubes have been touted as a more energy-efficient replacement for conventional silicon transistors in future generations of microchips. However, owing to imperfections inherent in carbon-nanotube devices, it is difficult to integrate them into large-scale electronic circuits. As a result, progress on this technology has been sluggish. On page 526 of this issue, Shulaker *et al.*¹ report a promising advance in this field — the first carbon-nanotube computer.

Usually it takes a whole armada of engineers to design and fabricate a functional computer from scratch, so it is worth noting how this small research group has made a nanotube computer. The authors took a two-pronged approach. First, they built on their technical know-how and experience of growing and aligning arrays of carbon nanotubes on a substrate. They developed methods to disable, on the substrate, all metallic carbon nanotubes, which would jeopardize the desired semiconducting behaviour of the system. The resulting substrate surface was covered with a highly aligned array of semiconducting nanotubes. The researchers were then able to create perfectly working transistors using advanced transistor-layout design and lithographic techniques, with each transistor consisting of a parallel arrangement of several individual semiconducting nanotubes.

By properly interconnecting the nanotube transistors, the authors succeeded in forming arbitrary logic elements and circuits. The underlying logic of the devices is the same as the p-type metal-oxide-semiconductor (PMOS) logic that was used in the early days

of semiconductor transistors, before n-type metal-oxide-semiconductor (NMOS) logic took over around 1970. A PMOS transistor is switched on when a negative voltage is applied to its control (gate) electrode. By contrast, an NMOS transistor turns on if a positive voltage is applied to the gate.

The second aspect of Shulaker and colleagues' approach involved choosing the simplest-possible computer design — thus reducing the complexity of the hardware circuitry, and so the number of transistors, required to attain the computer's desired functionality. The authors opted for a computer that operates on only 1 bit of information and uses a single instruction; today's computers normally involve 32 or 64 bits and use many instructions. But it has been shown that any n-bit operation can be obtained by using multiple 1-bit operations, although at the expense of being more time consuming. So the authors' method does not compromise on generality.

The only instruction that the computer executes is the SUBNEG (subtract and branch if negative) command², which in this design can be implemented with only 20 nanotube transistors. SUBNEG takes the content of a first memory address, subtracts it from the content of a second memory address and stores the result in the second memory address. If the result of this subtraction is negative, it goes to a third memory address. Because the instruction contains this conditional statement, it guarantees Turing completeness — that is, it can make any calculation if the computer has enough memory available. In other words, the instruction enables a universal computer to be made². With this one instruction, Shulaker and colleagues' nanotube computer was able to run counting and integer-sorting algorithms concurrently.

1. Stuart-Smith, R. D. *et al. Nature* **501**, 539–542 (2013).
2. Mora, C. *et al. PLoS Biol.* **9**, e1000606 (2011).
3. Safi, K. *et al. Phil. Trans. R. Soc. B* **366**, 2536–2544 (2011).
4. Newbold, T., Butchart, S. H. M., Şekercioglu, Ç. H., Purves, D. W. & Scharlemann, J. P. W. *PLoS ONE* **7**, e44019 (2012).
5. Tittensor, D. P. *et al. Nature* **466**, 1098–1091 (2010).
6. Kiessling, W., Simpson, C. & Foote, M. *Science* **327**, 196–198 (2010).

In terms of performance, this computer is far from being competitive with current standards, but had this machine been made in 1955, it would have been. The use of PMOS-only logic limits the scalability of the approach because this logic requires that the smallest transistor differs in width from the largest transistor by more than a factor of 20 or so. What is more, PMOS logic consumes electric power at all times, because there is always current flowing in the underlying circuitry. Today's silicon-based computer microchips operate with complementary metal-oxide-semiconductor (CMOS) technology, which uses PMOS and NMOS transistors of almost equal width in a serial connection. This allows CMOS logic to be considerably more scalable and to consume less power than PMOS or NMOS logic.

Implementing CMOS logic in carbon-nanotube circuitry is straightforward^{3,4}. For Shulaker *et al.*, CMOS implementation would just have doubled the number of steps in the computer's fabrication. However, the fabrication yield (the number of functional transistors) would have dropped. This is due to the fact that every step has a certain probability of creating defects in the devices. Consequently, if the number of steps increases, the probability of getting non-working devices increases. But the history of chip manufacturing has shown that yield increase is basically a matter of effort, and so there is no roadblock to engineering nanotube-based circuits in CMOS design.

The smallest transistor width used by Shulaker and colleagues is roughly 8 micrometres, owing to the statistical nature of the authors' nanotube-growth process. This leaves open the question of the ultimate scalability of their method, and so of the potential to bring it on a par with, or ahead of, current silicon technologies. The answer will depend on how precisely nanotubes can be arranged on a substrate. Fortunately, advances on this front have not stopped⁵, and a density of 500 nanotubes per micrometre might be feasible in the near future⁶. Therefore, if research efforts are focused towards delivering a scaled-up (64 bits) and scaled-down (20-nanometre transistor size) version of Shulaker and colleagues' nanotube computer, we might be able to type on one soon. ■

Franz Kreupl is in the Department of Hybrid Electronic Systems, Technische Universität München, 80333 München, Germany.
e-mail: franz.kreupl@tum.de

1. Shulaker, M. M. *et al.* *Nature* **501**, 526–530 (2013).
2. Gilreath, W. F. & Laplante, P. A. *Computer Architecture: A Minimalist Perspective* (Springer, 2003).
3. Chen, C., Xu, D., Kong, E. S.-W. & Zhang, Y. *IEEE Electron Dev. Lett.* **27**, 852–855 (2006).
4. Wang, C., Ryu, K., Badmaev, A., Zhang, J. & Zhou, C. *ACS Nano* **5**, 1147–1153 (2011).
5. Franklin, A. D. *Nature* **498**, 443–444 (2013).
6. Cao, Q. *et al.* *Nature Nanotechnol.* **8**, 180–186 (2013).

EARTH SCIENCE

A resolution of the Archaean paradox

A model of early Earth, in which heat from the interior is carried to the surface through volcanic heat pipes, suggests that our planet 4 billion years ago had more in common with Jupiter's moon Io than with today's Earth. [SEE ARTICLE P.501](#)

LOUIS MORESI

Earth formed 4.5 billion years ago from collisions between proto-planetary fragments. At that time, enough heat was trapped to melt much of Earth's interior; the segregation of the dense metallic core released still more heat; and a rich concentration of radioactive elements would have ensured that the interior stayed hot. This was the Earth of the Hadean eon, and only individual crystals are left of the rocks from that time. The first whole rocks in the geological record date from the Archaean eon, which started around 4 billion years ago. Despite all expectations of a hot young Earth, these first rocks reveal that the deep continental crust of the Archaean was no hotter than that of today. This puzzling observation is often known as the Archaean paradox. Reporting on page 501 of this issue, Moore and Webb¹ offer a resolution to it.

The research duo suggests that, during the first half of the Archaean, approximately 4 billion to 3.5 billion years ago, hot magma erupted quickly to the surface through heat pipes, so that the heat largely bypassed the crust. The surface became covered by one lava flow after another, with each flow rapidly cooling and gradually sinking into the interior as the surface layers built up. The authors argue that resurfacing the Earth, one lava flow at a time, at a rate of just 1–2 millimetres per year, would have been enough to extract the excess heat so efficiently that plate tectonics would have been completely suppressed for at least the first billion years of Earth's existence. About 3.5 billion years ago, Earth became cooler and plate tectonics took over.

Plate tectonics describes a relatively orderly turning inside-out of the planet: new ocean floor forms at narrow mid-ocean ridges and moves as a rigid plate towards the deep

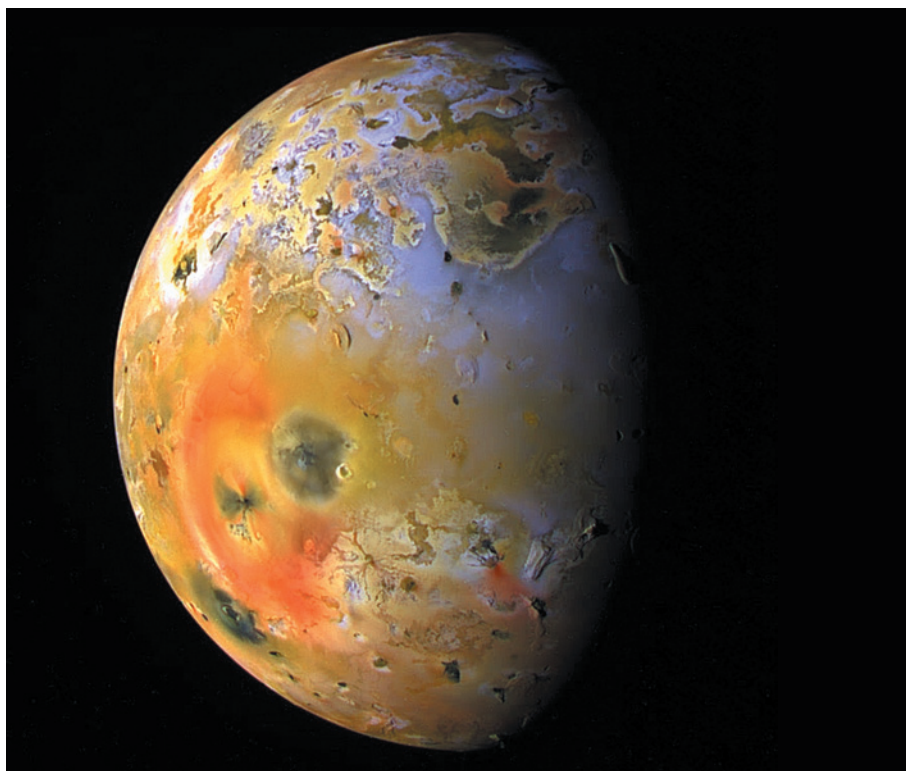


Figure 1 | Jupiter's moon Io. Moore and Webb's study¹ indicates that the early Earth's surface may have been similar to that of Jupiter's moon Io.

UNIVERSAL HISTORY ARCHIVE/UGC VIA GETTY IMAGES

trenches at the ocean margins, where it returns to Earth's deep interior in a process known as subduction. The continents, although they drift and deform in response to the plate movements, are largely spectators in the recycling of the surface. The continental geological record dates back all the way to the Archaean, whereas almost all the oceanic record that is more than 200 million years old has been subducted and lost. Plate tectonics explains how Earth works today, but has it always been this way? Did plate tectonics begin only in the mid-Archaean, as Moore and Webb's model requires? Looking to the continental geological record for answers, we find plenty of room for debate.

Certainly, many have argued that the geology of the early Archaean Earth is reminiscent of current conditions on Venus or on Jupiter's moon Io. These are dominated by volcanic processes from below, with little horizontal motion^{2,3}, in agreement with the authors' model (Fig. 1). Others have claimed, from a theoretical point of view, that plate tectonics works very differently in a hot Earth and may not be viable because hotter oceanic plates are too buoyant to subduct⁴, or because stresses in a hot Earth are much lower and it is less likely that the surface could be broken up into plates⁵.

Proponents of early plate tectonics during the first half of the Archaean point out that

the Archaean paradox is quite easy to resolve if the oceanic plates simply recycle faster in a hotter Earth^{6,7}. There are magmatic rocks from the Archaean that are similar to those found in present-day subduction zones, as well as structures imaged in the ancient crust that look as though they could have been created by subduction. Furthermore, recent theoretical models paint a more optimistic picture of the viability of subduction in a hotter planet⁸.

Moore and Webb's work is sure to re-energize this debate because they use the Archaean paradox to argue against early plate tectonics and their theoretical models produce plates in the late Archaean. Their simple model is elegant because it contains a pre-plate, heat-pipe Earth that undergoes a predictable transition to a plate-tectonic Earth as a consequence of cooling through time. In the study of Earth-like extrasolar planets, there is a fierce dispute over whether plate tectonics is a common phenomenon or unique to our planet⁹. Quantifying how plate tectonics can evolve on planets of different sizes and compositions is an important contribution to this wider debate.

In their model, which so far is limited to small-scale, two-dimensional, flat-Earth simulations, the transition from a pre-plate, heat-pipe Earth to a plate-tectonic one is

sudden and global. In reality, Earth is large enough for regional differences in the internal temperature to have been quite common even during major plate reorganization events¹⁰, so it might be possible for the two modes to coexist for some time. It will also be interesting to see whether the wild fluctuations in the global rate of volcanic eruption that are present in simple, flat-Earth models also occur in more realistic models. ■

Louis Moresi is in the School of Mathematical Sciences and the School of Geosciences, Monash University, Clayton, Victoria 3800, Australia.

e-mail: louis.moresi@monash.edu

1. Moore, W. B. & Webb, A. G. *Nature* **501**, 501–505 (2013).
2. Van Kranendonk, M. J., Smithies, R. H., Hickman, A. H. & Champion, D. C. *Terra Nova* **19**, 1–38 (2007).
3. O'Reilly, T. C. & Davies, G. F. *Geophys. Res. Lett.* **8**, 313–316 (1981).
4. Davies, G. F. *Lithos* **30**, 281–289 (1993).
5. O'Neill, C., Lenardic, A., Moresi, L., Torsvik, T. H. & Lee, C.-T. A. *Earth Planet. Sci. Lett.* **262**, 552–562 (2007).
6. Lenardic, A. *Geophys. J. Int.* **134**, 706–720 (1998).
7. Burke, K. & Kidd, W. S. F. *Nature* **272**, 240–241 (1978).
8. van Hunen, J. & Mosen, J.-F. *Annu. Rev. Earth Planet. Sci.* **40**, 195–219 (2012).
9. van Heck, H. J. & Tackley, P. J. *Earth Planet. Sci. Lett.* **310**, 252–261 (2011).
10. Lenardic, A. et al. *Geochim. Geophys. Geosyst.* **12**, Q10016 (2011).

CONDENSED-MATTER PHYSICS

Rotating molecules as quantum magnets

The push to engineer and probe quantum many-body systems using ultracold gases has reached a milestone with the observation of controlled dynamics caused by interactions between distant molecules trapped in a lattice. [SEE LETTER P.521](#)

ANDREW J. DALEY

Experiments with ultracold gases of atoms and molecules open up many avenues for exploring fundamental questions of many-body physics, because they offer a high level of control over the trapping potentials that confine the gases and interparticle interactions¹. On page 521 of this issue, Yan and colleagues² provide a highly anticipated demonstration of a new system in this context, with potassium–rubidium molecules trapped in separate sites of an optical lattice formed by standing waves of laser light*. They observe the dynamics caused by dipolar interactions between distant molecules, setting the stage for the exploration

of many-body phenomena and quantum phases of matter, especially those described by quantum spin models^{3–6}.

In Yan and colleagues' experiment, gases of potassium and rubidium at nanokelvin temperatures are associated to form K–Rb molecules by means of a magnetic-field ramp, and are then placed in the lowest-energy state of their electronic, vibrational and rotational degrees of freedom using laser-induced transitions. Ultracold gases of such heteronuclear ground-state molecules were first produced in 2008 (ref. 7), and are particularly interesting for the study of many-body physics because of the strong electric dipole moments exhibited by the molecules (0.57 debye units for K–Rb)^{3,6}.

But the authors have now gone a step further, introducing an optical lattice that traps

molecules at individual lattice sites (Fig. 1), separated by 0.5 micrometres, in which the lifetime of the molecular gas (limited by heating processes) is 25 seconds. This allows the realization of models in which the translational motion of the molecules is frozen and the dynamics are dominated by the molecules' internal motion. The molecules can each be treated as effective spin-1/2 systems by identifying two different states of their rotational motion as spin up and spin down. Applying microwave fields to couple the rotational states allows the dipole–dipole molecular interaction to exchange angular momentum between spatially separated molecules, giving rise to effective interactions between the spins.

Yan *et al.* measured the effects of these interactions on dynamics using a Ramsey spectroscopy technique, in which the rotation of the molecules is controlled through microwave pulses. The interactions show up as oscillations of the system's collective spin state, with a dominant oscillation frequency that is characteristic of the dipole–dipole interactions between molecules in neighbouring lattice sites. The authors also demonstrate how the effects of pairwise interactions can be cancelled using multiple microwave pulses, leaving a general damping of oscillations in the collective spin state that is caused by interactions between many molecules.

These measurements highlight the two key advantages of loading molecules into an

*This article and the paper under discussion² were published online on 18 September 2013.

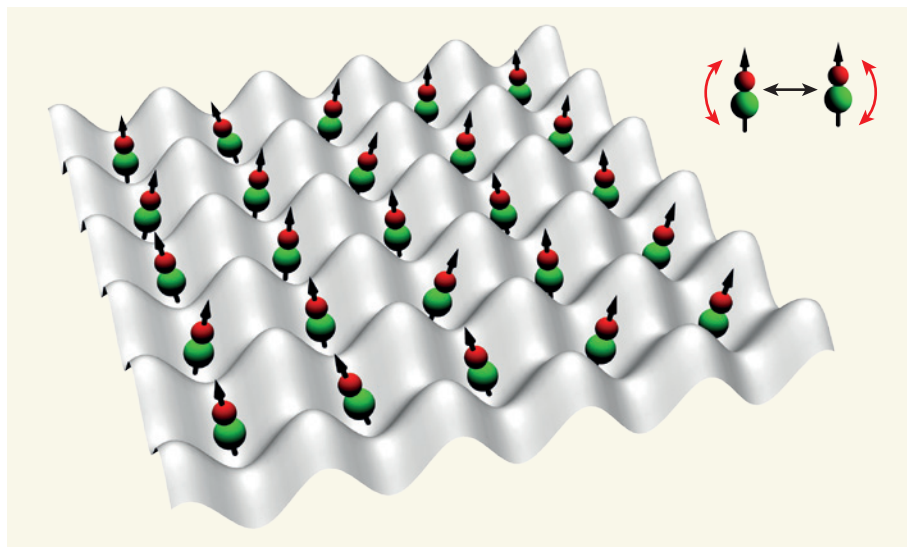


Figure 1 | Trapped in a lattice. Yan *et al.*² demonstrate a system of ultracold heteronuclear molecules (red and green) in an optical lattice (grey) that can be used to study quantum spin models. The lattice freezes the translational motion of the molecules, and a dipole–dipole interaction between molecules (dipoles shown as thick black arrows) couples different states of their rotational motion (red arrows), which can be used to represent quantum-mechanical spins. The authors' system currently has a smaller filling fraction of molecules than is shown here.

optical lattice. First, because the translational motion is frozen, the effective temperature in the rotational motion — and therefore the spin systems — can be very low, even if the initial temperature of the gas would have been too high to observe the dynamics cleanly. Second, trapping molecules in individual sites prevents collisions between molecules. In the case of K–Rb, collisions would lead to chemical reactions that produce K₂ and Rb₂ dimers⁸. This reaction releases a lot of energy, normally leading to the loss of molecules from the system.

Yan and colleagues also showed that when the molecules are allowed to move through the lattice, tunnelling weakly from site to site, their loss is further suppressed by a quantum-mechanical phenomenon known as the continuous quantum Zeno effect. In this mechanism, a dissipative process (the reactive loss) suppresses a coherent one (in this case, the tunnelling), so that the rapid loss of molecules from the same site actually decreases the likelihood that a molecule will move to a lattice site where another molecule is already present. Thus, counter-intuitively, a rapid loss process slows down the loss of molecules. Such suppression has previously been observed⁹ for highly excited molecules of Rb₂, but it has now been observed cleanly for the first time in ground-state molecules.

This experimental system has been anticipated by many theoretical proposals, which have demonstrated that many classes of spin model can be realized by tailoring the dipole–dipole interactions with external electric and microwave fields^{3–6}. This includes spin models that describe topological quantum phases⁴. Such exotic phases do not fit into the standard Landau theory that characterizes phase

transitions with a local-order parameter — they are instead identified by highly non-local topological properties. As shown by Yan *et al.*, these systems are particularly suitable for the study of non-equilibrium dynamics in strongly interacting spin models. Although some spin models can be implemented with atoms in optical lattices¹, the interactions in those systems are limited to contact interactions when the atoms collide, and effective spin–spin interactions involve tunnelling of particles in the lattice. Dipolar molecular systems such as those of Yan and colleagues provide a means to create spin models with stronger interactions,

allowing phenomena to be observed at much higher temperatures than those required in atomic systems.

At present, the biggest challenge in the K–Rb experiment is to increase the ‘filling fraction’ (the proportion of lattice sites occupied by a molecule), which is currently around 10%. This should ideally be close to 100% to realize some of the most interesting many-body models, which will require the production of molecular gases with higher initial density and at a lower temperature than achieved here. In the meantime, however, Yan and colleagues’ work immediately opens the door to the study of dynamics in disordered spin models, with the randomness arising from the distribution of molecules in the lattice. As technical improvements are made over the next few years, this system will provide an exciting path for the realization of exotic many-body physics. ■

Andrew J. Daley is in the Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA.

e-mail: adaley@pitt.edu

1. Bloch, I., Dalibard, J. & Nascimbène, S. *Nature Phys.* **8**, 267–276 (2012).
2. Yan, B. *et al. Nature* **501**, 521–525 (2013).
3. Carr, L. D., DeMille, D., Krevs, R. V. & Ye, J. *New J. Phys.* **11**, 055049 (2009).
4. Micheli, A., Brennen, G. K. & Zoller, P. *Nature Phys.* **2**, 341–347 (2006).
5. Barnett, R., Petrov, D., Lukin, M. & Demler, E. *Phys. Rev. Lett.* **96**, 190401 (2006).
6. Pupillo, G., Micheli, A., Büchler, H. P. & Zoller, P. in *Cold Molecules: Theory, Experiment, Applications* (eds Krevs, R. V., Stwalley, W. C. & Friedrich, B.) 421–470 (CRC Press, 2009).
7. Ni, K. K. *et al. Science* **322**, 231–235 (2008).
8. Ospelkaus, S. *et al. Science* **327**, 853–857 (2010).
9. Syassen, N. *et al. Science* **320**, 1329–1331 (2008).

CELL BIOLOGY

A table for two

Autophagy, the process of cellular self-cannibalism, comes in various forms. It now emerges that two of these — mitophagy and xenophagy — share a common initiator protein, Parkin. [SEE ARTICLE P.512](#)

MARCEL A. BEHR & ERWIN SCHURR

Infectious disease is not the inevitable consequence of exposure to a pathogen. Host factors have a crucial role in determining the outcome of such exposure, yet much remains unknown about how individuals vary in their capacity to resist pathogens. In this regard, genomic studies offer a tenable approach to identifying pathways involved in microbial handling. An intriguing example of how genomic findings can guide the mechanistic understanding of complex

biological systems is Manzanillo and colleagues’ paper¹ on page 512 of this issue. The study concerns the mechanism by which cells remove the human pathogen *Mycobacterium tuberculosis* through a process called autophagy*.

Genomic studies provide lists of genes, both expected and unexpected. Whereas expected genes serve to reinforce existing models, it is the set of genes with no apparent link to mechanism that is the most daunting. For instance,

*This article and the paper under discussion¹ were published online on 4 September 2013.

how can one explain the observation that *PARK2*, a gene first identified in autosomal recessive early-onset Parkinson's disease, has been identified in genomic studies^{2,3} as being associated with susceptibility to the infectious diseases leprosy and typhoid fever? Manzanillo *et al.* provide a functional link. They show that the product of *PARK2* — an E3 ubiquitin ligase enzyme called Parkin — has a role in xenophagy, an autophagic pathway that delivers intracellular bacteria for degradation in cellular organelles called lysosomes, and that is related to but distinct from mitophagy, an autophagic pathway that clears damaged cellular organelles known as mitochondria.

Mitochondria are a main source of cellular energy, which is generated in a process that depends on the presence of an electrochemical potential across the inner mitochondrial membrane. Mitochondria that fail to maintain this membrane potential are replaced as part of mitochondrial quality control⁴. Specifically, the kinase enzyme PINK1 senses dissipation of the inner-membrane potential and accumulates on the outer membrane of dysfunctional mitochondria (Fig. 1a). There, PINK1 is recognized by Parkin, which tags target mitochondrial proteins with the small protein ubiquitin for subsequent degradation by mitophagy⁵. At least in neurons, Parkin translocation to mitochondria also requires reactive oxygen species (ROS)⁶.

Mutated Parkin from patients with autosomal recessive Parkinson's disease cannot facilitate mitophagy⁵, suggesting that defective mitophagy contributes to neuronal death and neurodegeneration in this disorder. Unexpectedly, Manzanillo *et al.* find that Parkin is also involved in the clearance of microbial pathogens, but in a manner that is independent of its mitochondrial function.

The authors investigated the role of Parkin in mouse and human macrophages — immune cells that specialize in engulfing extracellular material such as pathogens by a process known as phagocytosis. To trigger the assembly of the autophagy complex, Parkin ubiquitinates phagosomal vesicles containing *M. tuberculosis*; the identity of the phagosomal proteins to which ubiquitin binds through its lysine amino-acid residue at position 63 is unknown. Consistent with an essential role for Parkin in delivering *M. tuberculosis* to lysosomes, deletion of *Park2* results in increased bacterial proliferation in infected macrophages in mice, and shorter survival times for these animals.

Autophagy has a key role in the destruction of invading bacteria such as mycobacteria and salmonella⁷.

Manzanillo and colleagues' work adds Parkin recruitment as an essential event in initiating the clearance of these intracellular pathogens by autophagy (Fig. 1b). Intriguingly, the authors find that Parkin mediates the control of infection with different intracellular pathogens (mycobacteria, salmonella and listeria) and in different hosts (mice and fruitflies). This suggests an evolutionarily conserved role for Parkin in innate immunity. Yet, despite its broad spectrum of pathogen control, Parkin seems to act on a different cellular compartment from that involved in the direct ubiquitination of cytosolic bacteria that is mediated by LRSAM1 — another E3 ligase involved in the autophagic destruction of *Salmonella* bacteria⁸.

Phagocytosis of salmonella by macrophages results in recruitment of the enzyme NOX2 NADPH oxidase to the phagosome

membrane and the production of ROS⁹. Generation of ROS is necessary for recruitment of the autophagy machinery to phagosomes and delivery of their bacterial cargo to the lysosome⁹. However, the factors that link ROS generation to autophagosome formation are not known. It is reasonable to propose that Parkin may be one such factor, given this enzyme's key role in both xenophagy and mitophagy^{1,5} and the requirement for ROS in both autophagy of salmonella-containing phagosomes and Parkin-dependent mitophagy in neurons.

The present paper raises further questions. For example, following infection of macrophages with *M. tuberculosis*, approximately one-third of the ingested bacteria are tagged by Parkin for autophagy. What is the fate of the two-thirds of bacteria that are not tagged, and do they benefit through a form of bacterial altruism? What is the membranous structure surrounding the bacteria that is tagged by Parkin for autophagy? Is it derived from the cell membrane as would be expected, or has it undergone structural and compositional changes to enable Parkin binding? Do bacterium-specific signals contribute to the autophagic response? Active Parkin ubiquitinates many proteins on damaged mitochondria¹⁰, and it will be interesting to contrast these with the unknown targets of phagosomal Parkin.

The capacity for functionally dissecting genomic findings has advanced tremendously since the completion of microbial and human genomes a decade or so ago. Manzanillo and colleagues' work reminds us that the unexpected results of genomic studies are often the most exciting. ■

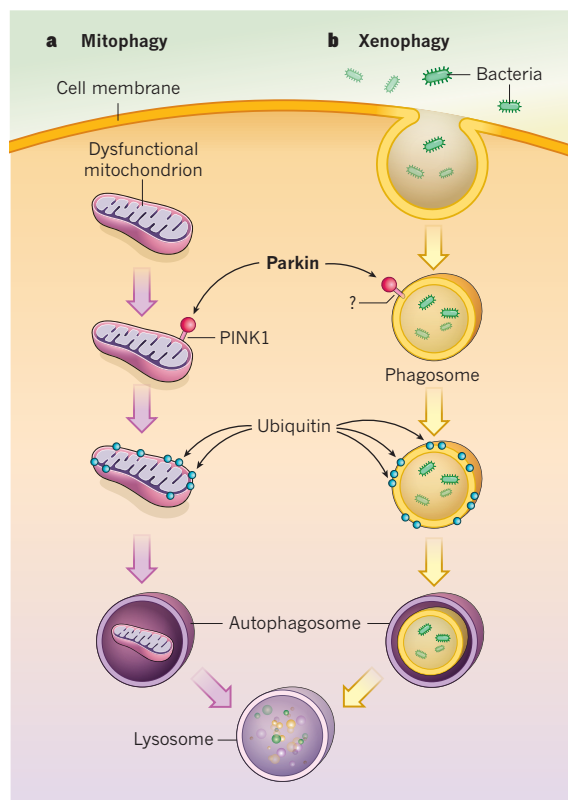


Figure 1 | Parkin and autophagy. **a**, The autophagic process of mitophagy is initiated by recruitment of the enzyme Parkin to dysfunctional mitochondria; this is mediated by both the enzyme PINK1 and reactive oxygen species (ROS; not shown). Subsequent activation of Parkin leads to addition of the small molecule ubiquitin to mitochondrial proteins and autophagy of the ubiquitinated mitochondria. This involves the engulfment of these organelles in autophagosomes and the fusion of autophagosomes with lysosomes, where the engulfed mitochondria are destroyed. **b**, Manzanillo *et al.*¹ find that Parkin is the link between mitophagy and another autophagic process, xenophagy. Xenophagy also depends on ROS and the binding of Parkin to phagosomes containing ingested intracellular pathogens (such as the bacteria *Mycobacterium tuberculosis* and *Salmonella Typhimurium*). How Parkin binds to phagosomes is not known. But once bound, Parkin ubiquitinates certain proteins, targeting the phagosome for lysosomal destruction. For clarity, only one of the two double membranes of the autophagosome is shown.

Marcel A. Behr and Erwin Schurr
are at the McGill International TB Centre, the Research Institute of the McGill University Health Centre and the Departments of Human Genetics, of Epidemiology & Biostatistics and of Medicine, McGill University, Montreal, Quebec H3G 1A4, Canada.
e-mails: marcel.behr@mcgill.ca; erwin.schurr@mcgill.ca

- Manzanillo, P. S. *et al.* *Nature* **501**, 512–516 (2013).
- Mira, M. T. *et al.* *Nature* **427**, 636–640 (2004).
- Ali, S. *et al.* *Clin. Exp. Immunol.* **144**, 425–431 (2006).
- Vincow, E. S. *et al.* *Proc. Natl Acad. Sci. USA* **110**, 6400–6405 (2013).
- Narendra, D. P. *et al.* *PLoS Biol.* **8**, e1000298 (2010).
- Joselin, A. P. *et al.* *Hum. Mol. Genet.* **21**, 4888–4903 (2012).
- Gutierrez, M. G. *et al.* *Cell* **119**, 753–766 (2004).
- Huett, A. *et al.* *Cell Host Microbe* **12**, 778–790 (2012).
- Huang, J. *et al.* *Proc. Natl Acad. Sci. USA* **106**, 6226–6231 (2009).
- Sarraf, S. A. *et al.* *Nature* **496**, 372–376 (2013).

Heat-pipe Earth

William B. Moore^{1,2} & A. Alexander G. Webb³

The heat transport and lithospheric dynamics of early Earth are currently explained by plate tectonic and vertical tectonic models, but these do not offer a global synthesis consistent with the geologic record. Here we use numerical simulations and comparison with the geologic record to explore a heat-pipe model in which volcanism dominates surface heat transport. These simulations indicate that a cold and thick lithosphere developed as a result of frequent volcanic eruptions that advected surface materials downwards. Declining heat sources over time led to an abrupt transition to plate tectonics. Consistent with model predictions, the geologic record shows rapid volcanic resurfacing, contractional deformation, a low geothermal gradient across the bulk of the lithosphere and a rapid decrease in heat-pipe volcanism after initiation of plate tectonics. The heat-pipe Earth model therefore offers a coherent geodynamic framework in which to explore the evolution of our planet before the onset of plate tectonics.

The lithospheric dynamics of terrestrial planets is driven by the transport of heat from the interior to the surface¹. Terrestrial bodies with low heat flows (for example Mars ($<20 \text{ mW m}^{-2}$; ref. 2) and the Moon (12 mW m^{-2} ; ref. 3)) lose heat largely by conduction across a single-plate lithosphere, whereas Earth's heat transport (global mean flux¹, 65 mW m^{-2}) is dominated by plate tectonics. Early in Earth's history, radiogenic heat production was three to five times greater than at present⁴, and there were additional contributions from tidal heating by a receding Moon and loss of accretionary heat. Whether or not plate tectonics operates under these conditions is uncertain geodynamically^{5–9}, but plate tectonic processes such as subduction and arc accretion are often invoked to explain the geologic and geochemical features of Archaean rocks^{10–12}.

An example of a terrestrial body with a higher surface heat flow than modern Earth is Jupiter's moon Io. Rather than losing heat by more vigorous plate tectonics, Io instead transports about 40 times Earth's heat flux¹³ (2.5 W m^{-2}) from the interior to the surface through volcanism. This mode of planetary heat transport is called the heat-pipe mode^{14,15} after the localized channels through which melt brings heat to the surface. Heat pipes are conduits that transfer heat and material from the base of the lithosphere to the surface by means of buoyant ascent (for example the lithospheric plumbing atop a mantle plume). When heat pipes become the dominant heat transport mechanism of a planet, the effects on the lithosphere are both surprising and profound.

Geodynamic models of heat-pipe Earth

We explore the consequences of the heat-pipe mode for early Earth using simplified models of mantle convection with melt generation and extraction. The temperature field snapshots shown in Fig. 1 result from numerical solutions to the equations of mass, energy and momentum transport in the mantle¹⁶ as internal heating and cooling at the surface drive convective motions (for details of the modelling approach and parameter definitions, see Methods Summary). These two-dimensional models have a strongly temperature-dependent, Newtonian rheology that results in a single-plate, rigid lid at the surface. Melting and melt transport are modelled in as simple a fashion as possible while preserving the effect of this process on the heat transport and dynamics of the lithosphere. Melt is generated whenever the mantle exceeds a simple linearly pressure-dependent solidus¹⁷ and is immediately extracted

to the surface, and the column in which the melt was produced is advected downwards to conserve mass (Fig. 1, inset). Once at the surface, the melt is assumed to lose its latent and sensible heat (which is tracked as volcanic heat flow) instantly and to return to the imposed surface boundary temperature of 15°C . Although this is a simplification of the process of melt generation and eruption, it captures the basic physics in a manner that allows us to see clearly the effects of heat pipes on the lithosphere.

The internal heating and surface heat flow both increase by a factor of ten from top to bottom in Fig. 1. Increased internal heat production causes the temperature of the mantle to increase slightly and the thickness of the cold lid to increase significantly. A thick, cold lithosphere develops because volcanic material deposited at the surface buries old flows, resulting in a descending 'conveyor belt' of material

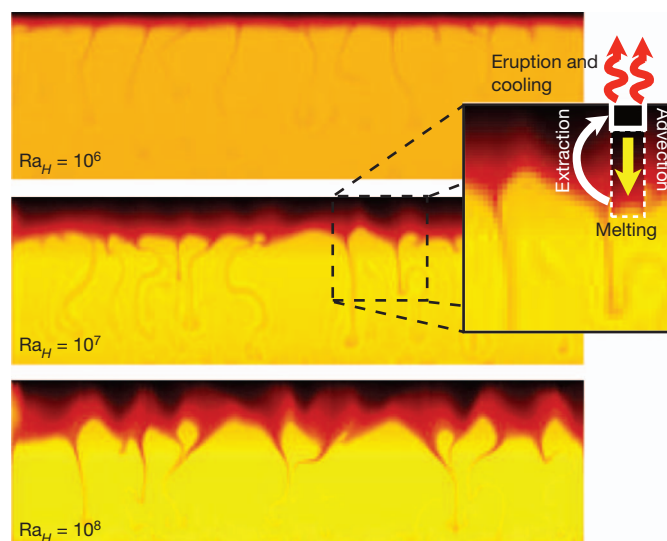


Figure 1 | Snapshots of the temperature field for two-dimensional models of mantle convection. The internal-heating Rayleigh number, Ra_H , is different in each panel. The inset illustrates the operation of the heat pipe: melt is extracting to the surface, where it cools, and cold lithosphere is advected downwards to conserve mass.

¹Department of Atmospheric and Planetary Sciences, Hampton University, Hampton, Virginia 23668, USA. ²National Institute of Aerospace, Hampton, Virginia 23666, USA. ³Department of Geology and Geophysics, Louisiana State University, Baton Rouge, Louisiana 70803, USA.

that advects the cold surface temperature to great depths (Fig. 1, inset). Because heat-pipe volcanism is global, the constant resurfacing and downward advection causes compression as the surface rocks are forced radially inward, resulting in uplift, exhumation and shortening. This process of contraction is observed on Io¹⁸, which has many mountains more than 10 km high that are clearly fault-bounded tectonic uplifts.

We explore the effect of heat pipes on the thermal structure of the lithosphere in more detail by constructing steady-state solutions of the energy equation with a term that accounts for the downward advection at a velocity v (ref. 14):

$$\frac{k}{\rho c_p} \frac{d^2 T}{dz^2} = v \frac{dT}{dz} - \frac{H}{\rho c_p} \quad (1)$$

where k is the thermal conductivity ($4 \text{ W m}^{-1} \text{ K}^{-1}$), ρ is the density ($3,000 \text{ kg m}^{-3}$), c_p is the specific heat capacity at constant pressure ($1,000 \text{ J kg}^{-1} \text{ K}^{-1}$) and H is the volumetric heat production rate (10^{-7} W m^{-3} at 4 Gyr ago). Material reaching the base of the lithosphere, defined by the dry peridotite solidus¹⁷, melts and returns to the surface through heat pipes. In Fig. 2, we plot solutions for three different lithospheric thicknesses (120, 150 and 180 km) and three different advection rates (0.1, 1 and 10 mm yr^{-1}). Rapid downward advection ($>0.5 \text{ mm yr}^{-1}$) brings materials to high pressures before they are significantly heated. Thermal equilibrium through heat pipes in early Earth requires a global average resurfacing by flows of $1\text{--}2 \text{ mm yr}^{-1}$ (whereas the rate on Io is 10 mm yr^{-1}). Partial melting and intra-lithospheric differentiation may also occur, resulting in more evolved lithologies erupting and intruding back up through the stack. The intrusions would locally modify the temperature structure by delivering heat to shallower regions. When advection slows or stops, the lithosphere re-warms as geotherms relax towards the conductive profile.

In addition to altering the thickness and thermal structure of the lithosphere, heat pipes influence the stress distribution as shown in Fig. 3, where we plot the maximum stress observed in the lithosphere as a function of the internal heating rate. Volcanic heat pipes are the dominant heat transport mechanism at high heat production rates (Fig. 3a, solid line), reflecting higher mantle temperatures and increased rates of melting. The maximum stress experienced in the lithosphere (Fig. 3b) has a distinct minimum that coincides with the crossover between heat-pipe and conductive heat transport. This behaviour occurs because heat pipes remove buoyancy from the actively convecting boundary layer at the top of the mantle, reducing convective stresses, while at the same time producing a thicker lithosphere. Therefore, active

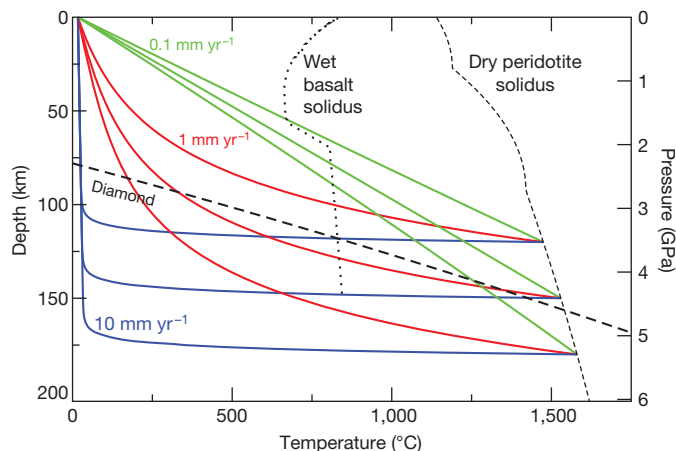


Figure 2 | Temperature as a function of depth in lithosphere. Three downward advection rates (0.1 , 1 and 10 mm yr^{-1}) are shown, each for three lithosphere thicknesses (120, 150 and 180 km). Also shown are the stability limit of diamond³⁹ (thick dashed), the wet basalt solidus¹¹ (dotted) and the dry peridotite solidus¹⁷ (thin dashed), which is chosen to define the base of the lithosphere.

heat pipes would lead to a very different lithospheric stress regime in early Earth than at present, as indicated in Fig. 3b.

We also performed simulations to identify the transition out of the heat-pipe mode of heat transport (Fig. 4). For these cases, we modelled the finite strength of the lithosphere by using a yield stress (Methods Summary). The transition is caused by the decline in heat sources (red line), which gradually causes the volcanic heat flow (cyan line) to decrease. At about the time the volcanic heat flow becomes small compared with the conductive heat flow (black line), the plate breaks and the entire lithosphere is replaced in a sudden overturn. This is an artefact of our two-dimensional approach; in three dimensions, the overturn would replace only portions of the lithosphere. Breaking occurs because the peak lithospheric stress increases with declining heat production, eventually overcoming the strength of the lithosphere and allowing the negative buoyancy of the cold material to drive rapid flow. There is a rapid decrease in volcanism (cyan) after lithospheric overturn.

Geologic evidence for heat-pipe Earth

The heat-pipe Earth hypothesis and simulations supply predictions for early Earth's development: rapid volcanic resurfacing, a low geothermal gradient across the bulk of the lithosphere, contractional deformation (with minor extensional deformation in restricted settings such as grabens atop rising diapirs) and a rapid decrease in heat-pipe volcanism after initiation of plate tectonics. Preserved heat-pipe rocks would be broadly representative of the single-plate lithosphere. Here we review the geologic record of early Earth and compare it with these predictions and those of existing models. We focus on the period before ~ 3.2 Gyr ago because interpretations of diverse data suggest that plate tectonics began at that time: isotopic records in zircons indicate a pronounced decrease in crustal growth rate¹⁹, inclusion assemblages of kimberlite-hosted diamonds shift from exclusively peridotitic to eclogitic as well as peridotitic²⁰, and a Wilson cycle that occurred 3.2–3.1 Gyr ago has been identified²¹.

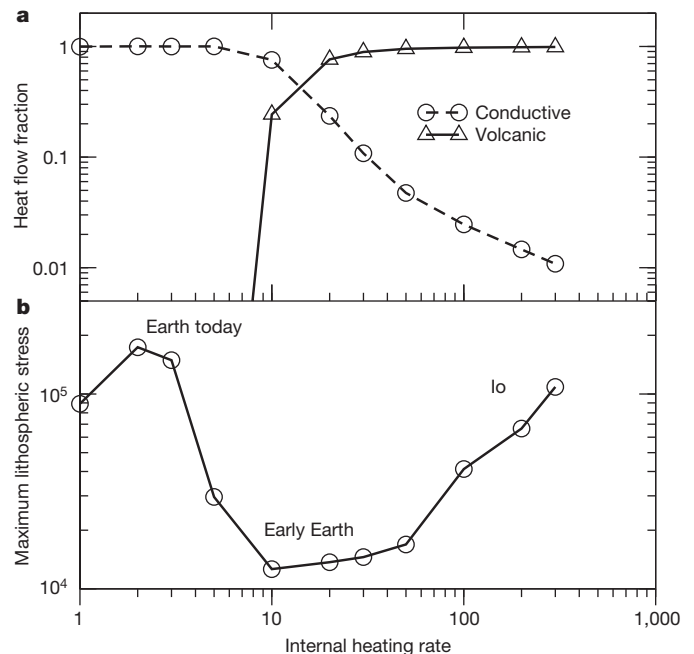


Figure 3 | Heat flow contributions and dimensionless maximum lithospheric stress as functions of the dimensionless internal heating rate ($\alpha HD^2/k$). a, Heat flow; b, maximum lithospheric stress. Volcanic heat-pipe transport dominates at higher internal heating rates. This causes a marked decrease in lithospheric stress. At very high heating rates (Io), lithospheric thickness variations lead to larger stresses. All quantities are long-term averages over quasi-steady conditions.

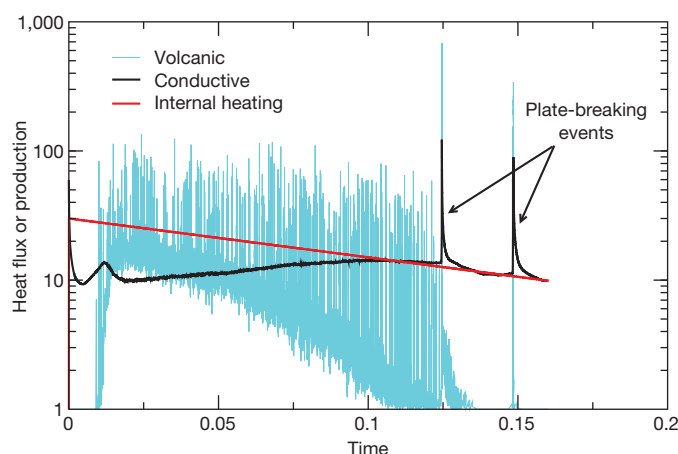


Figure 4 | Internal heat production, surface conductive heat transport and surface volcanic heat transport as functions of time in a model with a finite yield stress. The heat production has been converted into an equivalent surface flux by integrating over depth. The time is made dimensionless using the thermal diffusion timescale. Plate-breaking events are identified at times 0.125 and 0.15.

The best-preserved rock records from before 3.2 Gyr ago are from the Barberton (South Africa) and Pilbara (Australia) greenstone belts, which extend back to 3.55 and 3.53 Gyr ago, respectively^{22,23}. Before ~3.2 Gyr ago, both belts experienced similar geologic histories. Mafic and ultramafic volcanic rocks were erupted in alternating subaerial and marine settings and occur as thick sequences: ~10 km thick at Barberton and ~20 km at Pilbara^{22–24}. Minor chemical sediments, locally derived clastic sediments and felsic tuffs are interspersed in the mafic piles and commonly occur below unconformities^{22,23,25}. Deposition of volcanic and sedimentary rocks was episodic and semiregular. For example, at Pilbara eight cycles, each ~15 Myr long, have been identified, with the five deepest cycles forming a >12-km-thick section deposited over 100 Myr (ref. 25). Tonalite–trondhjemite–granodiorite (TTG) plutons comprise about half of the >3.2-Gyr-old rocks exposed across both sites and were intruded in several 30–60-Myr-long episodes spanning the 300-Myr history of each belt^{23,26}. The TTG rocks represent melts of two sources: pre-existing TTG and hydrated mafic rocks²⁶. Melts of hydrated mafic rocks were generated at pressures and temperatures ranging from 1.5 GPa and ~1,000 °C to ≥3.0 GPa and ~1,200 °C (refs 11, 26), suggesting heat flows of <50 to 80 mW m^{–2}. The upper limit of the pressure estimates is poorly constrained because the melt reactions have little to no pressure dependence¹¹. The more deeply sourced melts show evidence of interaction with ultramafic material during melt ascent²⁶.

Emplacement of the TTG plutons as rising diapirs developed a dome-and-syncline regional structural pattern, with the volcanic sequences deformed into downwelled synclinal troughs^{25,27}. Internal deformation associated with this downwelling may thin the volcanic sequences by 50% or more²⁷, causing estimated eruption rates to be lower limits. Apart from the diapir deformation, both belts are essentially undeformed until ~3.2 Gyr ago. At this time, rifting, arc development and accretion occurred at Pilbara²¹, whereas deformation and metamorphism associated with either collisional tectonics²² or rapid diapir emplacement²⁷ occurred at Barberton. These are the last medium–high-temperature events experienced by most portions of both belts^{22,28}.

Other than at Barberton and Pilbara, the >3.2 Gyr-old rock record is limited to gneiss complexes as old as ~4.03 Gyr ago²⁹. Some of the complexes show >3.2-Gyr-old migmatization, suggesting deformation before that time³⁰. The 3.85–3.55-Gyr-old Itsaq gneiss complex (Greenland) is broadly representative of these gneisses and preserves the best-defined deformation record. It experienced high-grade metamorphism, granite production and terrane juxtaposition along a series of mylonitic shear zones over a 100-Myr period^{10,30} (~3.65–3.55 Gyr

ago). In addition to granite, gneiss protolith lithologies here include all Barberton and Pilbara lithologies as well as limited slices of ≥3.65-Gyr-old peridotite^{10,31}. More than 95% of the rocks are metamorphosed felsic plutons with igneous crystallization ages spanning the age of the complex^{10,30}. Two suites of felsic rocks are derived from hydrated mafic rocks both above and below ~45-km depth, and from remelted TTG^{10,32}. Like at Barberton and Pilbara, maximum melting depths are unconstrained¹¹. Clastic metasediments are dominantly derived from the mafic or ultramafic volcanics, with minor detrital input from the other lithologies³³. All detrital material is restricted to narrow age ranges just slightly older than the timing of sedimentation³³.

Although the rock record ends at ~4 Gyr ago, detrital zircon grains provide evidence of earlier times: the oldest zircons from the Jack Hills locality (Australia) are nearly 4.4 Gyr old. Hafnium, oxygen and lead isotope information from zircon grains have been used to demonstrate that a population of Jack Hills zircon grains ranging in age from ~4.3 to 3.9 Gyr were sourced from melts of weathered material that formed in the presence of water^{34,35}, that a population of Jack Hills zircon grains grew in a ~4.1-Gyr-old felsic pluton which itself resulted from remelting of 4.36–4.30-Gyr-old TTG³⁶, and that the sources of TTGs have not changed during the past 4.3 Gyr (ref. 12). Also, a population of Jack Hills zircon grains ranging in age from ~4.25 to 3.2 Gyr features variable Th/U ratios and inclusions of ultrahigh-pressure metamorphic diamonds³⁷. The diamonds show a wide range of carbon isotope ratios, even in the same zircon, indicating a diversity of carbon sources that most probably reflects surface processes³⁸. Therefore, the diamonds probably represent former surface material, and they crystallized at depths of >100 km and were subsequently included in igneous zircon^{37,38}.

Comparison of the >3.2-Gyr-old geologic record with the predictions of heat-pipe Earth reveals that a wide range of structural, kinematic, petrologic and isotopic data are consistent with the operation of heat pipes. The long duration of rapid volcanic resurfacing observed at Barberton and Pilbara matches the main surface prediction of the heat-pipe hypothesis. The thickness and age limits of the volcanic sequences allow estimation of downward advection rates: approximately 0.03, 0.07 and 0.12 mm yr^{–1} for the Barberton, Pilbara and lower-Pilbara sections, respectively. These rates must be considered minima because the preserved volcanic sequences contain unconformities and occur in the thinned limbs of downwelled synclines. Similarly, the restricted age range of clastic sediment source material at Itsaq might indicate rapid volcanic resurfacing³³. In general, we speculate that interpretations applied to the Barberton and Pilbara greenstone belts may be similarly applicable to Itsaq because of the broad correspondence in protolith lithologies.

Downward advection accounts for the low geothermal gradients implied by melting conditions. Felsic volcanics and TTG plutons may all be sourced from a downwards-advecting lithosphere comprising TTG and mafic or ultramafic materials. Melting of the shallow lithosphere to produce low-pressure TTG melts probably occurs in response to thermal perturbations caused by the rise of hot, deeply sourced TTG melts. The possible rapid TTG diapir emplacement and the onset of rifting ~3.2 Gyr ago are consistent with the rewarming of the lithosphere associated with both the end of heat-pipe cooling and the initiation of plate tectonics. Heat-pipe cooling would have to be significantly diminished after ~3.2 Gyr ago, because the Barberton and Pilbara rocks did not experience significant subsequent burial. This is consistent with the abrupt drop in volcanism seen in Fig. 4.

Within the context of the heat-pipe hypothesis, deformation such as the ~3.65–3.55-Gyr-old protracted tectonothermal event at Greenland might be interpreted as either development of a subduction zone¹⁰ or a long-lived reverse-fault duplex system. The first option may be possible because our simulations suggest that heat-pipe mode would continue to contribute to cooling after a subduction event (Fig. 4). Nonetheless, the second option seems more likely given the abrupt decrease in heat-pipe volcanism after ~3.2 Gyr ago. Also, peridotite has been interpreted

as oceanic crust in the first hypothesis^{10,31}, but these rocks may represent ultramafic plutonic rocks or deeply buried lavas.

Jack Hills zircon constraints demonstrate that TTGs as old as ~4.4 Gyr were generated by melting and remelting of material that was previously weathered at the surface^{12,34–36}. This is consistent with the ~4.0–3.2-Gyr-old rock record, suggesting that analogous geodynamic interpretations apply at least as far back as ~4.4 Gyr ago. The inclusion of diamond with likely surface signatures in similar zircons is interpreted in the heat-pipe Earth context to indicate that the lithospheric conveyor belt was operating to a depth of >100 km ~4.25 Gyr ago, and plausibly until 3.2 Gyr ago, which is consistent with the cold, thick lithosphere produced in heat-pipe Earth (Figs 1 and 2). Also shown in Fig. 2 are melting curves for wet basalt¹¹ and dry peridotite¹⁷, and the phase boundary for diamond³⁹. Because the only means of exhuming these diamonds is the ascent of TTG plutons, some of these plutons must be sourced within the diamond stability field. This is consistent with the intersection of the wet basalt solidus and the heat-pipe geotherm (for example the red lines in Fig. 2).

Previous explanations of the >3.2-Gyr-old geologic record can be loosely divided into protoplate tectonic models and vertical tectonics models⁴⁰. Plate tectonic models involve horizontal plate motion, with spreading ridges and subduction zones to create and reprocess lithosphere⁴¹. Vertical tectonic models posit lithospheric evolution through sub- and intra-lithospheric diapirism, associated downwelling and volcanism, and basal delamination. The central challenge of the early-Earth record for plate tectonic models is that important products of plate tectonics, many of which have high preservation potential (for example paired metamorphic belts and passive margins), are absent⁴⁰. In contrast, the record contains features not readily explained by proposed vertical tectonic processes. Namely, terrane juxtaposition such as seen at Greenland, where peridotite slices may be interpreted as ophiolites, has been explained only by incorporating horizontal motion akin to plate motion^{40,42}. As a result of including such kinematics, these models face similar problems as plate tectonic models.

The geology of Barberton and Pilbara illustrates the difficulties inherent in previous models. These sites preserve no deformation for >300 Myr, with the exception of diapiric deformation. Current plate and vertical tectonic models explain this geology in terms of prolonged mantle plume volcanism^{23,27}. However, all such models involve significant horizontal translation of lithosphere, which presents two problems. First, it would be highly unusual in a plate tectonic regime for two plumes, or two successions of plumes, to remain fixed at both sites for >300 Myr. Second, there is no evidence of horizontal motion in the record. Cooling by means of heat pipes does not have similar problems because it features semi-continuous volcanism and minor horizontal contraction globally.

A new model for early Earth tectonics

The heat-pipe hypothesis implies that early Earth's lithospheric dynamics was very different from that today. Similar to Jupiter's volcanically active moon Io, early Earth transported heat in the heat-pipe mode, with melt rising through narrow conduits and downward advection of the lithosphere beneath newer flows. The lithosphere is thickened and the geotherm is depressed by the advection (Figs 1 and 2). Lithospheric stresses are reduced by a combination of thickening and the loss, through the heat pipes, of buoyancy from the convective thermal boundary layer at the top of the mantle (Fig. 3). These model predictions are all consistent with the main features of the >3.2-Gyr-old geologic record, which is dominated by deeply sourced, high-melt-fraction volcanic rocks that have interacted with water and sedimentary products from the surface. The transition from heat pipes to plate tectonics has also been modelled, showing a rapid switch between modes (Fig. 4) that matches the geologic record at Barberton and Pilbara. Heat-pipe Earth provides a coherent dynamic framework for the understanding of planetary evolution from initial crystallization of the lithosphere atop the magma ocean to the onset of plate tectonics.

METHODS SUMMARY

The calculations presented here use the STAG3D code¹⁶ to solve the equations of mass, momentum and energy conservation for infinite-Prandtl-number Boussinesq flow in a 4×1 , two-dimensional domain spanning the depth of the mantle. The system is made dimensionless by using the mantle depth, D , as the length scale, D^2/κ as the time scale (κ , thermal diffusivity) and $D^2 H/k$ as the temperature scale, resulting in the definition¹ of the internal-heating Rayleigh number $Ra_H = \alpha \rho g H D^5 / k \eta \kappa$ with thermal expansivity α , gravitational acceleration g and viscosity η . For all simulations presented here, $Ra_H/H = 10^6$.

The Newtonian, exponentially temperature-dependent rheology has a total viscosity contrast that ranges from 10^4 to 10^7 as H varies from 1 to 300. The computational grid is 256×64 or 512×128 depending on the Rayleigh number.

For the time-dependent simulation with a finite yield stress, $H = 30$ and an effective viscosity, η_{eff} , was specified to incorporate a yield stress σ_y as follows¹⁶:

$$\sigma_{ij} = 2 \frac{\sigma_y^2 \eta}{\sigma_y^2 + \eta^2 \dot{\epsilon}^2} \dot{\epsilon}_{ij} \equiv 2 \eta_{\text{eff}} \dot{\epsilon}_{ij}$$

Here σ_{ij} and $\dot{\epsilon}_{ij}$ are the stress and strain rate components, respectively, and $\dot{\epsilon}^2$ is the square of the second invariant of the strain tensor. The value of the yield stress was chosen to intersect the maximum lithospheric stress curve in the lower panel of Fig. 3.

The one-dimensional models of equation (1) have constant temperature boundary conditions applied at the surface and at an arbitrary depth identified as the base of the lithosphere.

Received 24 March; accepted 12 July 2013.

1. Turcotte, D. L. & Schubert, G. *Geodynamics* 1–17, 2nd edn (Cambridge Univ. Press, 2002).
2. McGovern, P. J. *et al.* Correction to “Localized gravity/topography admittance and correlation spectra on Mars: implications for regional and global evolution”. *J. Geophys. Res.* **109**, E07007 (2004).
3. Warren, P. H. & Rasmussen, K. L. Megaregolith insulation, internal temperatures, and bulk Uranium content of the Moon. *J. Geophys. Res.* **92**, 3453–3465 (1987).
4. Franck, S. Evolution of the global mean heat flow over 4.6 Gyr. *Tectonophysics* **291**, 9–18 (1998).
5. Davies, G. F. Gravitational depletion of the early Earth's upper mantle and the viability of early plate tectonics. *Earth Planet. Sci. Lett.* **243**, 376–382 (2006).
6. Korenaga, J. Thermal evolution with a hydrating mantle and the initiation of plate tectonics in the early Earth. *J. Geophys. Res. Solid Earth* **116**, B12403 (2011).
7. O'Neill, C. J., Lenardic, A., Moresi, L., Torsvik, T. H. & Lee, C.-T. A. Episodic pre-Cambrian subduction. *Earth Planet. Sci. Lett.* **262**, 552–562 (2007).
8. van Hunen, J. & van den Berg, A. P. Plate tectonics on the early Earth: limitations imposed by strength and buoyancy of subducted lithosphere. *Lithos* **103**, 217–235 (2008).
9. van Hunen, J. & Moya, J.-F. Archean subduction: fact or fiction? *Annu. Rev. Earth Planet. Sci.* **40**, 195–219 (2012).
10. Nutman, A. P., Friend, C. R. L., Horie, K. & Hikada, H. in *Earth's Oldest Rocks* (eds Van Kranendonk, M. J., Smithies, R. H. & Bennett, V. C.) 187–218 (Elsevier, 2007).
11. Moya, J.-F. The composite Archean grey gneisses: petrological significance, and evidence for a non-unique tectonic setting for Archean crustal growth. *Lithos* **123**, 21–36 (2011).
12. Guitreau, M., Blichert-Toft, J., Martin, H., Mojszisz, S. J. & Albareda, F. Hafnium isotope evidence from Archean granitic rocks for deep-mantle origin of continental crust. *Earth Planet. Sci. Lett.* **337–338**, 211–223 (2012).
13. Veeder, G. J., Matson, G. J., Johnson, T. V., Davies, A. G. & Blaney, D. L. The polar contribution to the heat flow of Io. *Icarus* **169**, 264–270 (2004).
14. O'Reilly, T. C. & Davies, G. F. Magma transport of heat on Io: a mechanism allowing a thick lithosphere. *Geophys. Res. Lett.* **8**, 313–316 (1981).
15. Breuer, D. & Moore, W. B. in *Treatise on Geophysics* Vol. 10 (ed. Schubert, G.) 299–341 (Elsevier, 2007).
16. Tackley, P. J. Modelling compressible mantle convection with large viscosity contrasts in a three-dimensional spherical shell using the yin-yang grid. *Phys. Earth Planet. Inter.* **171**, 7–18 (2008).
17. Takahashi, E. Melting of a dry peridotite KLB-1 up to 14 GPa: implications on the origin of peridotitic upper mantle. *J. Geophys. Res.* **91**, 9367–9385 (1986).
18. Schenk, P. M. & Bulmer, M. H. Origin of mountains on Io by thrust faulting and large-scale mass movements. *Science* **279**, 1514–1517 (1998).
19. Dhuime, B., Hawkesworth, C. J., Cawood, P. A. & Storey, C. D. A change in the geodynamics of continental growth 3 billion years ago. *Science* **335**, 1334–1336 (2012).
20. Shirey, S. B. & Richardson, S. H. Start of the Wilson cycle at 3 Ga shown by diamonds from subcontinental mantle. *Science* **333**, 434–436 (2011).
21. Pease, V., Percival, J., Smithies, R. H., Stevens, G. & Van Kranendonk, M. in *When Did Plate Tectonics Begin On Planet Earth?* (eds Condie, K. C. & Pease, V.) 199–228 (The Geological Society of America, 2008).
22. Lowe, D. R. & Byerly, G. R. in *Earth's Oldest Rocks* (eds Van Kranendonk, M. J., Smithies, R. H. & Bennett, V. C.) 481–526 (Elsevier, 2007).
23. Van Kranendonk, M. J., Smithies, R. H., Hickman, A. H. & Champion, D. C. in *Earth's Oldest Rocks* (eds Van Kranendonk, M. J., Smithies, R. H. & Bennett, V. C.) 307–337 (Elsevier, 2007).

24. Stiegler, M. T., Lowe, D. R. & Byerly, G. R. Fragmentation and dispersal of komatiitic pyroclasts in the 3.5–3.2 Ga Onverwacht Group, Barberton greenstone belt, South Africa. *Geol. Soc. Am. Bull.* **123**, 1112–1126 (2011).
25. Hickman, A. H. & Van Kranendonk, M. J. in *The Precambrian Earth: Tempos and Events* (eds Eriksson, P. G., Altermann, W., Nelson, D. R., Mueller, W. U. & Catuneanu, O.) 54–75 (Elsevier, 2004).
26. Moya, J.-F., Stevens, G., Kisters, A. F. M. & Belcher, R. W. in *Earth's Oldest Rocks* (eds Van Kranendonk, M. J., Smithies, R. H. & Bennett, V. C.) 607–667 (Elsevier, 2007).
27. Van Kranendonk, M. J. Cool greenstone drips and the role of partial convective overturn in Barberton greenstone belt evolution. *J. Afr. Earth Sci.* **60**, 346–352 (2011).
28. Zegers, T. E., Wijbrans, J. R. & White, S. H. $^{40}\text{Ar}/^{39}\text{Ar}$ age constraints on tectonothermal events in the Shaw area of the eastern Pilbara granite-greenstone terrain (W. Australia): 700 Ma of Archaean tectonic evolution. *Tectonophysics* **311**, 45–81 (1999).
29. Bowring, S. A. & Williams, I. S. Priscoan (4.00–4.03 Ga) orthogneisses from northwestern Canada. *Contrib. Mineral. Petrol.* **134**, 3–16 (1999).
30. Friend, C. R. L. & Nutman, A. P. Complex 3670–3500 Ma orogenic episodes superimposed on juvenile crust accreted between 3850–3690 Ma, Itsaq Gneiss Complex, southern West Greenland. *J. Geol.* **113**, 375–397 (2005).
31. Friend, C. R. L., Bennett, V. C. & Nutman, A. P. Abyssal peridotites >3,800 Ma from southern West Greenland: field relationships, petrology, geochronology, whole-rock and mineral chemistry of dunite and harzburgite inclusions in the Itsaq Gneiss Complex. *Contrib. Mineral. Petrol.* **143**, 71–92 (2002).
32. Hoffmann, J. E. *et al.* Mechanisms of Archean crust formation inferred from high-precision HFSE systematics in TTGs. *Geochim. Cosmochim. Acta* **75**, 4157–4178 (2011).
33. Kamber, B. S., Whitehouse, M. J., Bolhar, R. & Moorbath, S. Volcanic resurfacing and the early terrestrial crust: zircon U-Pb and REE constraints from the Isua Greenstone Belt, southern West Greenland. *Earth Planet. Sci. Lett.* **240**, 276–290 (2005).
34. Mojzsis, S. J., Harrison, T. M. & Pidgeon, R. T. Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* **409**, 178–181 (2001).
35. Wilde, S. A., Valley, J. W., Peck, W. H. & Graham, C. M. Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* **409**, 175–178 (2001).
36. Blichert-Toft, J. & Alverède, F. Hafnium isotopes in Jack Hills zircons and the formation of the Hadean crust. *Earth Planet. Sci. Lett.* **265**, 686–702 (2008).
37. Menneken, M., Nemchin, A. A., Geisler, T., Pidgeon, R. T. & Wilde, S. A. Hadean diamonds in zircon from Jack Hills, Western Australia. *Nature* **448**, 917–920 (2007).
38. Nemchin, A. A. *et al.* A light carbon reservoir recorded in zircon-hosted diamond from the Jack Hills. *Nature* **454**, 92–95 (2008).
39. Bundy, F. P. Pressure-temperature phase-diagram of elemental carbon. *Physica A* **156**, 169–178 (1989).
40. Stern, R. J. Modern-style plate tectonics began in Neoproterozoic time: an alternative interpretation of Earth's tectonic history. *Spec. Pap. Geol. Soc. Am.* **440**, 265–280 (2008).
41. Harrison, T. M. The Hadean crust: evidence from >4 Ga zircons. *Annu. Rev. Earth Planet. Sci.* **37**, 479–505 (2009).
42. Bédard, J. H., Harris, L. B. & Thurston, P. C. The hunting of the snArc. *Precamb. Res.* **229**, 20–48 (2013).

Acknowledgements This research has been supported by NSF Geophysics, NASA PG&G, the NASA Astrobiology Institute and a start-up fund from Louisiana State University, and has made use of the Astrophysics Data System. We thank M. Stiegler for discussions.

Author Contributions Both authors contributed to the writing of this paper. W.B.M. performed the numerical modelling and A.A.G.W. provided the review of the geologic literature.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.B.M. (william.moore@hamptonu.edu).

Transcriptome and genome sequencing uncovers functional variation in humans

Tuuli Lappalainen^{1,2,3}, Michael Sammeth^{4,5,6,7,†*}, Marc R. Friedländer^{5,6,7,8*}, Peter A. C. 't Hoen^{9*}, Jean Monlong^{5,6,7*}, Manuel A. Rivas^{10*}, Mar González-Porta¹¹, Natalja Kurbatova¹¹, Thasso Griebel⁴, Pedro G. Ferreira^{5,6,7}, Matthias Barann¹², Thomas Wieland¹³, Liliana Greger¹¹, Maarten van Iterson⁹, Jonas Almlöf¹⁴, Paolo Ribeca⁴, Irina Pulyakhina⁹, Daniela Esser¹², Thomas Giger¹, Andrew Tikhonov¹¹, Marc Sultan¹⁵, Gabrielle Bertier^{5,6}, Daniel G. MacArthur^{16,17}, Monkol Lek^{16,17}, Esther Lizano^{5,6,7,8}, Henk P. J. Buermans^{9,18}, Ismael Padioulet^{1,2,3}, Thomas Schwarzmayer¹³, Olof Karlberg¹⁴, Halit Ongen^{1,2,3}, Helena Kilpinen^{1,2,3}, Sergi Beltran⁴, Marta Gut⁴, Katja Kahlem⁴, Vyacheslav Amstislavskiy¹⁵, Oliver Stegle¹¹, Matti Pirinen¹⁰, Stephen B. Montgomery[†], Peter Donnelly¹⁰, Mark I. McCarthy^{10,19}, Paul Flicek¹¹, Tim M. Strom^{13,20}, The Geuvadis Consortium[‡], Hans Lehrach^{15,2†}, Stefan Schreiber¹², Ralf Sudbrak^{15,2†}, Ángel Carracedo²², Stylianos E. Antonarakis^{1,2}, Robert Häsler¹², Ann-Christine Syvänen¹⁴, Gert-Jan van Ommen⁹, Alvis Brazma¹¹, Thomas Meitinger^{13,20,23}, Philip Rosenstiel¹², Roderic Guigó^{5,6,7}, Ivo G. Gut⁴, Xavier Estivill^{5,6,7,8} & Emmanouil T. Dermitzakis^{1,2,3}

Genome sequencing projects are discovering millions of genetic variants in humans, and interpretation of their functional effects is essential for understanding the genetic basis of variation in human traits. Here we report sequencing and deep analysis of messenger RNA and microRNA from lymphoblastoid cell lines of 462 individuals from the 1000 Genomes Project—the first uniformly processed high-throughput RNA-sequencing data from multiple human populations with high-quality genome sequences. We discover extremely widespread genetic variation affecting the regulation of most genes, with transcript structure and expression level variation being equally common but genetically largely independent. Our characterization of causal regulatory variation sheds light on the cellular mechanisms of regulatory and loss-of-function variation, and allows us to infer putative causal variants for dozens of disease-associated loci. Altogether, this study provides a deep understanding of the cellular mechanisms of transcriptome variation and of the landscape of functional variants in the human genome.

Interpreting functional consequences of millions of discovered genetic variants is one of the biggest challenges in human genomics¹. Although genome-wide association studies (GWAS) have linked genetic loci to various human phenotypes and the functional annotation of the genome is improving^{2,3}, we still have a limited understanding of the underlying causal variants and biological mechanisms. One approach to addressing this challenge has been to analyse variants affecting cellular phenotypes, such as gene expression^{4–8}, known to affect many human diseases and traits^{9,10}.

In this study, we characterize functional variation in human genomes by RNA-sequencing hundreds of samples from the 1000 Genomes Project¹, the most important reference data set of human genetic variation, thus creating the biggest RNA sequencing data set of multiple human populations so far. We not only catalogue novel loci with regulatory variation, but also, for the first time, discover and characterize molecular properties of causal functional variants.

We performed mRNA and small RNA sequencing on lymphoblastoid cell line samples from five populations: the CEPH (CEU), Finns (FIN),

British (GBR), Toscani (TSI) and Yoruba (YRI). After quality control, we had 462 and 452 individuals (89–95 per population) with mRNA and miRNA data, respectively (Supplementary Figs 1–11 and Supplementary Table 1). Of these, 421 are in the 1000 Genomes Phase 1 data set¹, and the remainder were imputed from single nucleotide polymorphism (SNP) array data (Supplementary Fig. 3 and Supplementary Table 2). High-throughput RNA sequencing (RNA-seq) was performed in seven laboratories, and the smaller amount of variation between laboratories than individuals demonstrated that RNA sequencing is a mature technology ready for distributed data production (Mann-Whitney $P < 2.2 \times 10^{-6}$ for mRNA, $P = 1.34 \times 10^{-10}$ for miRNA; Fig. 1a, Supplementary Fig. 11; for further details see ref. 11). To discover genetic regulatory variants, we mapped *cis*-quantitative trait loci (QTLs) to transcriptome traits of protein-coding and miRNA genes separately in the European (EUR) and Yoruba (YRI) populations (Table 1, Supplementary Fig. 12 and Supplementary Table 3). The RNA-seq read, quantification, genotype and QTL data are available open-access (see Author Information section).

¹Department of Genetic Medicine and Development, University of Geneva Medical School, 1211 Geneva, Switzerland. ²Institute for Genetics and Genomics in Geneva (iG3), University of Geneva, 1211 Geneva, Switzerland. ³Swiss Institute of Bioinformatics, 1211 Geneva, Switzerland. ⁴Centro Nacional de Análisis Genómico, 08028 Barcelona, Catalonia, Spain. ⁵Centre for Genomic Regulation (CRG), 08003 Barcelona, Catalonia, Spain. ⁶Pompeu Fabra University (UPF), 08003 Barcelona, Catalonia, Spain. ⁷CRG Hospital del Mar Research Institute, 08003 Barcelona, Catalonia, Spain. ⁸CRG CIBERESP, 08003 Barcelona, Catalonia, Spain. ⁹Department of Human Genetics, Leiden University Medical Center, 2300 RC Leiden, the Netherlands. ¹⁰Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. ¹¹European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK. ¹²Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, D-24105 Kiel, Germany. ¹³Institute of Human Genetics, Helmholtz Zentrum München, 85764 Neuherberg, Germany. ¹⁴Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, 751 85 Uppsala, Sweden. ¹⁵Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany. ¹⁶Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ¹⁷Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. ¹⁸Leiden Genome Technology Center, 2300 RC Leiden, the Netherlands. ¹⁹Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, Oxford OX3 7BN, UK. ²⁰Institute of Human Genetics, Technische Universität München, 81675 Munich, Germany. ²¹Dahlem Centre for Genome Research and Medical Systems Biology, 14195 Berlin, Germany. ²²Fundacion Publica Galega de Medicina Xenomica (SERGAS), Genomic Medicine Group, CIBERER, Universidade de Santiago de Compostela, Santiago de Compostela, Spain. ²³Deutsches Forschungszentrum für Herz-Kreislaufkrankungen (DZHK), Partner Site Munich Heart Alliance, 81675 Munich, Germany. [†]Present addresses: Bioinformatics Laboratory, National Laboratory of Scientific Computing (LNCC), Petropolis 25651-075, Rio de Janeiro, Brazil (M.S.); Departments of Pathology and Genetics, Stanford University, Stanford, California 94305-5324, USA (S.B.M.); Alacris Theranostics GmbH, 14195 Berlin, Germany (R.S.).

*These authors contributed equally to this work.

‡A list of authors and their affiliations appears in the Supplementary Information.

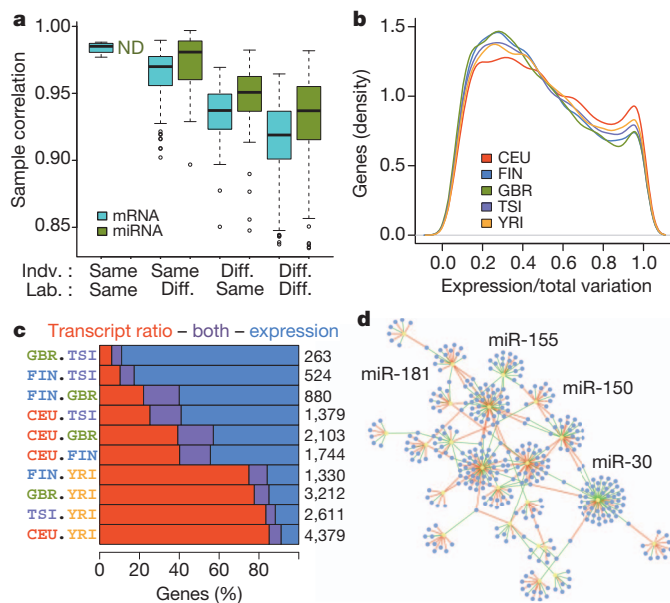


Figure 1 | Transcriptome variation. **a**, Spearman rank correlation of replicate samples, based on mRNA exon and miRNA quantifications of 5 individuals sequenced 8 and 7 times for mRNA and miRNA, respectively, and separated by the individual (indv.) or the sequencing laboratory (lab.) being the same or different (diff.). The quantifications have been normalized only for the total number of mapped reads (see Supplementary Fig. 11 for correlations after normalization). **b**, The proportion of expression level variation (as opposed to splicing) of the total transcription variation between individuals in each population, measured per gene. **c**, Proportion of genes with differential expression levels and/or transcript usage between population pairs, out of the total listed on the right-hand side. **d**, Network of significant miRNA families ($P < 0.001$; yellow) and their significantly associated mRNA targets ($P < 0.05$; purple). The edges display negative (green) and positive (red) associations.

Transcriptome variation in populations

This first uniformly processed RNA-seq data set from multiple human populations allowed high-resolution analysis of transcriptome variation. Individual and population differences in transcripts can manifest in (1) overall expression levels, and (2) relative abundance of transcripts from the same gene (transcript ratios). Deconvolution of the relative contribution of these¹² indicates that this ratio is characteristic for each gene, with transcript ratio being on average more dominant (Fig. 1b and Supplementary Figs 13 and 14). Population differences explain a small but significant proportion of 3% of the total variation (Mann-Whitney $P < 2.2 \times 10^{-16}$). In addition to this genome-wide perspective to population variation, we identified 263–4,379 genes with differential expression and/or transcript ratios between population pairs (P.G.F. *et al.* manuscript submitted). Notably, continental differences between YRI–EUR population pairs have a much higher contribution of genes with different transcript usage than European population pairs (75–85% versus 6–40%; Fig. 1c and Supplementary Fig. 14). This has not been observed before in humans, but it is consistent with splicing patterns capturing phylogenetic differences between species better than expression levels^{13,14}.

We quantify a total of 644 autosomal miRNAs in >50% individuals, of which 60 have significant *cis*-eQTLs for miRNA expression

levels (*cis*-mirQTLs, Table 1 and Supplementary Fig. 15), showing that genetic effects on miRNA expression are much more widespread than the previously identified loci¹⁵. To complement previous studies of miRNA function in cell perturbation experiments, we analysed miRNA–mRNA interactions in our steady-state population sample. Of 100 miRNA families, 32 correlated with the expression of predicted target exons in a highly connected network ($P < 0.001$; Fig. 1d and Supplementary Table 4), including miRNA families with important immunological or lymphocyte functions, such as miR-150, miR-155, miR-181 and miR-146 (ref. 16). Interestingly, 45% of the associations were positive—consistent with previous results¹⁵—even though based on perturbation experiments miRNAs mostly downregulate genes. Analysing the direction of causality, *cis*-mirQTLs had small *trans*-eQTL effects to predicted targets only when effects were negative ($\Pi_1 = 1 - \text{Storey's } \Pi_0 = 0.11$ versus $\Pi_1 = 0$; Supplementary Fig. 16), suggesting that miRNAs indeed downregulate their targets. Positive correlations may be driven by other effects, which is supported by overrepresentation of transcription factors in the network (29%, Fisher $P = 2.1 \times 10^{-7}$ for negative targets and 26% $P = 4.0 \times 10^{-4}$ for positive targets). This suggests feedback loops of both mRNA and miRNA genes affecting the expression of each other, and supports the idea that under steady-state conditions, miRNAs confer robustness to expression programs¹⁷. Altogether, these results highlight the added insight into the role of miRNAs in regulatory networks from analysis of population variation.

Genetic effects on the transcriptome

Expression QTL (eQTL) analysis of protein-coding and long inter-genic non-coding RNA (lincRNA) genes uncovered extremely widespread regulatory variation, with 3,773 genes having a classical eQTL for gene expression levels (Table 1). Although the potential of RNA-seq to discover other transcriptome traits such as splicing variation is widely known^{7,8,18–20}, a comprehensive analysis has been lacking. To this end, we first mapped eQTLs for exon quantifications that can capture both gene expression and splicing variation, discovering as many as 7,825 genes with an eQTL, referred to as eQTLs in this paper unless otherwise specified. Regressing out the most significantly associated variant from the EUR eQTL analysis showed that as many as 34% of the genes have a second, independent eQTL for any of their exons (of which 7% for the exon of the first association). Thus, there is substantial allelic heterogeneity for regulatory effects on a single gene and independence of exons of the same gene (Supplementary Fig. 17). To investigate genetic effects specifically on splicing, we discovered 639 genes with transcript ratio QTLs (trQTLs) affecting the ratio of each transcript to the gene total—the largest number of genetic effects on transcript structure identified so far. The lower number relative to gene eQTLs is probably caused by higher noise in model-based transcript quantifications than in gene counts. To characterize the relationship of genetic variants affecting expression versus splicing, we regressed out the best trQTL variant from the gene eQTL analysis in 279 genes with both types of QTL. The results showed that the causal variants are independent in $\geq 57\%$ of these genes (Supplementary Fig. 18), suggesting that transcriptional activity and transcript usage are usually controlled by different regulatory elements of the genome.

The transcript differences driven by trQTLs involve exon skipping only in 15% of genes, with as much as 48% and 43% varying in 5' and 3' ends, respectively (in EUR; categories not mutually exclusive; Fig. 2b). To analyse transcript modifications further through unannotated transcript elements, we mapped *cis*-eQTLs for expressed retrotransposon-derived elements (repeat elements) outside genes, known to be an important source for evolution of new transcripts²¹. We detected widespread sharing between the 5,763 *cis*-eQTLs discovered for repeat elements (Table 1 and Supplementary Fig. 19) and nearby exon eQTLs: of the best repeat eQTLs variants in EUR, 49% were significant and 6% the top eQTL variants for exons of a nearby gene (3.8 \times and 26 \times enrichment; Fisher $P < 2.2 \times 10^{-16}$). This suggests that retrotransposon-derived elements can share regulatory elements with nearby genes. These results provide

Table 1 | Numbers of transcriptome features with a QTL (FDR 5%)

	Total	EUR ($n = 373$)	YRI ($n = 89$)	Union
Exon eQTL	12,981 genes	7,390	2,369	7,825
Gene eQTL	13,703 genes	3,259	501	3,773
Transcript ratio QTL	7,855 genes	620	83	639
mirQTL	644 miRNAs	57	15	60
Transcribed repeat eQTL	43,875 repeats	5,763	1,055	6,069

FDR, false discovery rate.

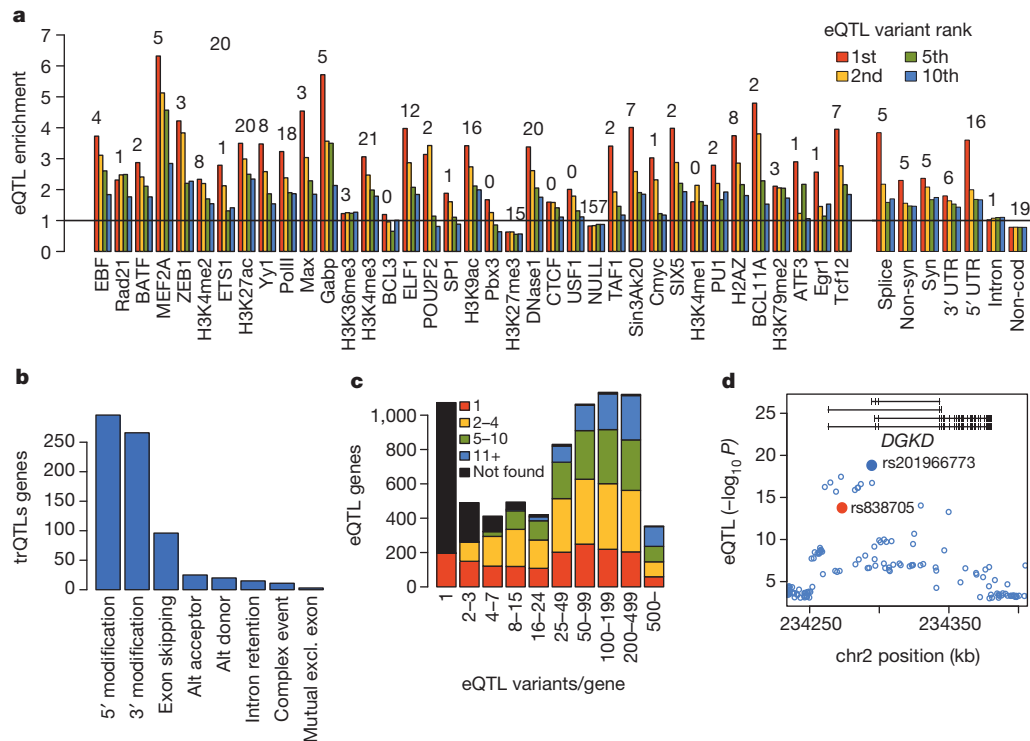


Figure 2 | Transcriptome QTLs. **a**, Enrichment of EUR exon eQTLs in functional annotations for the first, second, fifth and tenth best associating eQTL variant per gene, relative to a matched null set of variants denoted by the horizontal line. The numbers are $-\log_{10}(P)$ values of a Fisher test between the best eQTL and the null. UTR, untranslated region. **b**, Classification of changes

the first, to our knowledge, genome-wide characterization of genetic effects on transcript structure through annotated and unannotated 3' and 5' changes, which may predominate the exon skipping that previous studies have focused on¹⁹. This opens new perspectives for understanding their cellular and high-level effects, as end modifications will rarely change protein structure but may affect post-transcriptional regulation.

Altogether, we present the largest and the most diverse catalogue of *cis*-regulatory variants discovered in a single tissue so far. Most of the analysed genes—8,329 out of 13,970—have one or several QTLs for different transcript traits, a resolution enabled by in-depth analysis of high-quality transcriptome and genome sequencing data. These results highlight both allelic heterogeneity of regulatory variants and phenotypic heterogeneity of diverse transcriptome traits of individual genes.

Properties of regulatory variants

To understand how eQTLs affect gene expression, we compared the properties of the top (most significant) eQTL variant per gene to a null of non-eQTL variants (matched for distance from transcription start site (TSS) and minor allele frequency). The best eQTL variant may not always be the causal variant owing to noise in genotype and phenotype data, and to estimate our ability to pinpoint causal variants, we contrasted the properties of the first eQTL to the second, fifth and tenth best eQTL variants (Fig. 2a).

First, comparing the eQTL with the best *P* value to the matched null showed an enrichment of indels among top eQTLs (13% = $1.22 \times$ enrichment; Fisher $P = 1.9 \times 10^{-3}$ in EUR; Supplementary Fig. 20), suggesting that indels are more likely to have functional effects than SNPs. eQTLs are highly enriched in several non-coding elements from the Ensembl Regulatory Build, such as many transcription factor peaks (median enrichment $3.3 \times$, median $P = 0.009$ in EUR; Fig. 2a and Supplementary Fig. 21), DNase1 hypersensitive sites ($3.4 \times$, $P = 1.00 \times 10^{-20}$), as well as in chromatin states of active promoters ($3.5 \times$, $P = 1.08 \times 10^{-36}$) and strong enhancers (median $2.4 \times$, median $P = 1.14 \times 10^{-5}$). Within

caused by transcript ratio QTLs. **c**, The rank of the best Omni 2.5M SNP among the significant EUR eQTL variants per gene. **d**, The *DGKD* gene locus, in which an intronic SNP rs838705 is associated with calcium levels (red), and the top eQTL variant 21 kb downstream (blue) is a very likely causal variant, close to the TSS of two transcripts in the MEF2A, C binding region.

genes, splice-site ($3.8 \times$, $P = 1.65 \times 10^{-5}$) and non-synonymous ($2.3 \times$, $P = 4.84 \times 10^{-6}$) enrichments point to putative regulatory functions of coding variants. Transcript ratio QTLs are overrepresented in splice sites ($6.8 \times$, $P = 2.44 \times 10^{-7}$; Supplementary Fig. 22), as expected, but also, for example, in 3' untranslated regions ($2.5 \times$, $P = 1.83 \times 10^{-6}$) and promoters ($2.4 \times$, $P = 5.79 \times 10^{-6}$). Altogether, the higher resolution of annotations and eQTLs relative to previous studies^{22,23} provides important insight into the role of individual transcription factors and other regulatory elements mediating genetic regulatory effects.

Functional enrichment typically decreases rapidly from the best eQTL variant towards lower ranks. To estimate how often the first variant is likely to be the causal regulatory variant, we calculated the annotation enrichment of the best eQTL variants relative to the null for (1) all eQTL loci, and (2) loci in which the best eQTL variant is very likely causal owing to having a $\log_{10} P$ -value > 1.5 higher than the second variant (Supplementary Fig. 23). The ratio of the enrichments (1) and (2) yields an approximation of the best variant being causal in 55% of EUR and 74% of YRI eQTLs, with more conservative estimates being 34% and 41%, respectively (Supplementary Fig. 23). Thus, we have reasonable power to pinpoint causal regulatory variants from unbiased *P*-value distributions alone without annotation priors²³. This is enabled by not relying on SNP array data²²: in 81% of the cases the best variant is not on the Omni 2.5M array (Fig. 2c and Supplementary Fig. 25). Validating the putative causal effects, we observed that the best eQTL variants in CTCF peaks showed more allele-specific binding compared to matched null variants ($P = 2.0 \times 10^{-3}$; Supplementary Fig. 24) using CTCF ChIP-seq data from six individuals²⁴, and the best eQTLs were enriched in DNase1 hypersensitivity QTLs²⁵ ($3.3 \times$, $P = 2.51 \times 10^{-6}$ in EUR, $7.9 \times$, $P < 2.2 \times 10^{-16}$ in YRI). In conclusion, we not only identify broad eQTL loci but also substantially increase our confidence to pinpoint individual causal variants and their functional mechanisms.

Of the 6,473 variants in the GWAS catalogue²⁶, 16% are eQTLs and 1.8% are trQTLs in EUR or YRI, but a high overlap is observed also by

chance for a frequency-matched GWAS null (11% and 0.84%, respectively). The modest (albeit significant: $eQTL \chi^2 P < 2.2 \times 10^{-16}$; $trQTL P = 7.2 \times 10^{-9}$) enrichment^{9,10} is due to eQTLs being very ubiquitous, and consequently, a GWAS variant being an eQTL does not mean that the regulatory change is necessarily driving the disease association. Our data offers a unique opportunity to address the key question of whether the causal eQTL variant is also causal for the disease. The enrichment of GWAS SNPs in the top eQTL ranks ($P = 1.18 \times 10^{-7}$; Supplementary Fig. 26) is a genome-wide signal of shared causality. To characterize individual loci further, we selected 78 eQTL regions that are likely causal signals for 91 GWAS SNPs (estimated by the regulatory trait concordance method)^{6,9}, and in these loci our best eQTL variant is the putative disease-causing variant (Supplementary Fig. 27 and Supplementary Table 5). Figure 2d shows an example of the *DGKD* gene, in which an intronic SNP rs838705 is associated with calcium levels²⁷, and 21 kilobases (kb) downstream the top eQTL—a 2-base pair (bp) insertion—is the likely causal variant affecting calcium levels. Thus, the integration of genome sequencing and cellular phenotype data helps to not only understand causal genes and biological processes but also pinpoint putative causal genetic variants underlying GWAS associations.

Allelic and loss-of-function effects

Transcript differences between the two haplotypes of an individual allow quantification of regulatory variation even when eQTLs cannot be detected, for example, owing to low allele frequency. We analysed both allele-specific expression (ASE) and allele-specific transcript structure (ASTS), a novel approach based on exonic distribution of reads (Supplementary Figs 2 and 28–33). This first genome-wide quantification of allelic effects on transcript structure shows that it is almost equally common as ASE, with significant ($P < 0.005$) ASE and ASTS in a median of 6.5% and 5.6% sites (out of 8,420 and 2,135) per individual, respectively. Furthermore, the substantial overlap of ASE and ASTS signals (Fig. 3a) suggests that ASE may actually often be driven by transcript structure variation. The low population frequency of the vast majority of ASE (Fig. 3b) and ASTS (Supplementary Fig. 30) events points to widespread rare regulatory variation that is undetectable in eQTL analysis.

An important caveat in ASE analysis has been the possibility that it can be driven by purely epigenetic effects rather than *cis*-regulatory genetic variants. We investigated this by a novel approach to quantify concordance between ASE and putative regulatory variants (prSNPs), in which heterozygotes but not homozygotes for a true rSNP should have differential expression of the two haplotypes, that is, allelic imbalance in an aseSNP (Supplementary Figs 2 and 34). We calculated concordance of allelic ratios of 5,479 aseSNPs and genotypes of all variants ± 100 kb from TSS, with an empirical P value from 100–1,000 permutations. Assigning the prSNPs with empirical P -value < 0.01 to $P < 0.001$ as

likely rSNPs yielded a total of 224,640 rSNPs (7.4% of tested; Supplementary Table 6) that clustered close to TSS as expected for regulatory variants⁵ and replicated most eQTL signals (Supplementary Fig. 35). Nearly all aseSNPs (95%) had more observed rSNPs than expected; thus ASE seems to nearly always be genetic rather than driven by genotype-independent allelic epigenetic effects. rSNP signals are widespread and robust also outside eQTL genes (Supplementary Table 6 and Supplementary Fig. 35), indicating potential to capture novel effects. Variants that are both eQTLs and rSNPs show higher enrichment in functional annotations (Fig. 3c and Supplementary Fig. 36), suggesting that integrated analysis may improve resolution to find causal regulatory variants. Altogether, we show evidence that ASE effects are mostly rare and nearly always genetic, and ASE-based analyses may complement eQTL analysis in identification of especially low-frequency regulatory variants in future studies.

Although QTL and prSNP analyses aim at identifying previously unknown regulatory variants, we can also quantify functional effects of predicted loss-of-function variants²⁸. Our RNA-seq data captures 839 premature stop codon and 849 splice-site variants, with the much higher number than in previous studies enabling proper quantification of their transcriptome effects. As expected, premature stop variants often show loss of the variant allele (Supplementary Fig. 37), indicating nonsense-mediated decay²⁹ (NMD) as in previous studies^{28,30}. Variants close to the end of the transcript seem to escape NMD as predicted²⁹. However, of the variants predicted to trigger NMD, in 68% (54% of rare variants with minor allele frequency $< 1\%$) the ASE results do not support this (Fig. 4a), suggesting currently unknown mechanisms of NMD escape.

Finally, we modelled how genetic variants affect splicing affinity in the entire splicing motif rather than only the canonical splice site, which is the first comprehensive set of such predictions genome-wide (P.G.F. *et al.*, manuscript submitted). Non-reference alleles have a lower splicing affinity on average ($P < 2.2 \times 10^{-16}$; Supplementary Fig. 38). For the 10% of these variants predicted to destroy the motif, individuals carrying two motif-destroying alleles have 29% lower median inclusion rates of the affected exon ($P < 2.2 \times 10^{-16}$; Fig. 4b), indicating that our RNA-seq data are consistent with predictions of splicing effects.

Conclusions

By integrated analysis of RNA and DNA sequencing data we were able to obtain a unique view of variation of the transcriptome and its genetic causes, moving beyond eQTL catalogues to a high-resolution view of genetic regulatory variants. We deconvoluted the effect of gene expression and transcript structure in population differences of the transcriptome, in QTLs, and in allele-specific effects, and show that these two dimensions of transcript variation appear equally common but largely independent. Genetic regulatory variation is the rule rather than the

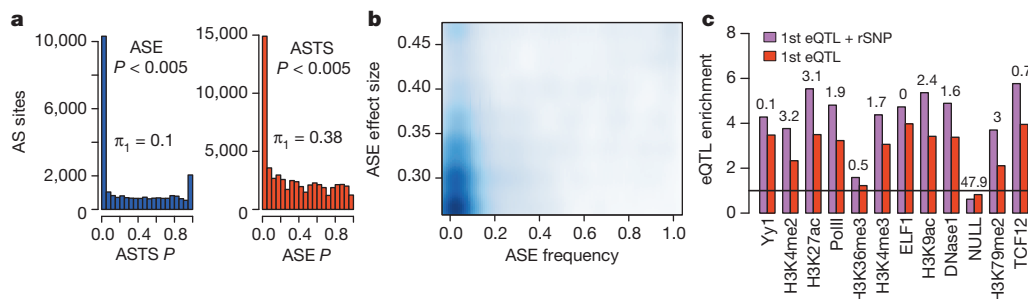


Figure 3 | Allele-specific effects on expression and transcript structure. **a**, Sharing of allele-specific expression (ASE) and transcript structure (ASTS) signals: the distribution of ASTS P value of the sites with significant ($P < 0.005$) ASE in the same individual, and vice versa. The ASE P -values are calculated from sites sampled to exactly 30 reads. The numbers denote the Π_1 (1 – Storey’s Π_0) statistic measuring the enrichment of low P values. **b**, Frequency of significant ASE event in the population (x axis) and their effect size

($|0.5 - \text{ref}/\text{total}|$), calculated per ASE SNP. Only ASE SNPs with ≥ 20 heterozygote individuals with ≥ 30 reads were included, and the estimates were corrected for coverage bias and false positives by sampling and permutations. **c**, Enrichment of variants in regulatory annotations relative to a matched null distribution for the most significant eQTL variants, and for the subset of these that are also rSNPs. Categories with highest amount of data are shown (see Supplementary Fig. 36 for all categories, see also Fig. 2a).

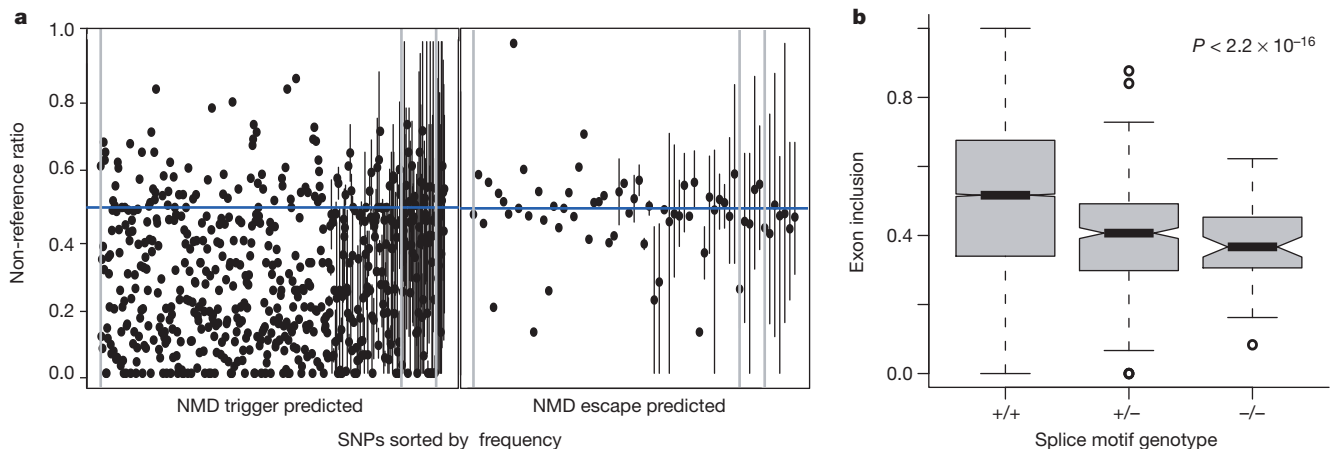


Figure 4 | Transcriptome effects of loss-of-function variants. **a**, NMD due to premature stop codon variants was measured using allele-specific expression. The distribution of non-reference allele ratios (on the y axis) for premature stop variants sorted on the x axis according to derived allele frequency, split by sites predicted to trigger and escape NMD. The dots denote the median across

exception in the genome with widespread allelic heterogeneity, and is the major determinant of allelic expression. For the first time, we were able to predict large numbers of causal regulatory variants, and thus provide a detailed view into cellular mechanisms of regulatory and loss-of-function variation, which is essential for future functional prediction of variants discovered in personal genomes.

A subset of this functional variation at the cellular level will also have effects on higher-level traits. We demonstrate how eQTL data can be used to pinpoint putative causal GWAS variants of individual loci, which is important as a new model of how integration of cellular phenotypes and genome sequencing data can uncover both causal variants and biological mechanisms underlying diseases. The landscape of regulatory variation in this study adds a functional dimension to the 1000 Genomes Project data, which is used in effectively all disease studies, and together they form an important joint reference data set of variation and function of the human genome. Ultimately, this study illustrates the power of combining genome sequence analysis with a high-depth functional readout such as the transcriptome.

METHODS SUMMARY

Total RNA was extracted from Epstein-Barr-virus-transformed lymphoblastoid cell line pellets using TRIzol reagent (Ambion), and mRNA and small RNA sequencing of 465 unique individuals were performed on the Illumina HiSeq2000 platform, with paired-end 75-bp mRNA-seq and single-end 36-bp small-RNA-seq. Five samples were sequenced in replicate in each of the seven sequencing laboratories. The mRNA and small RNA reads were mapped with GEM³¹ and miraligner³², respectively, with an average of 48.9 million mRNA-seq reads and 1.2 million miRNA reads per sample after quality control. Numerous transcript features were quantified using Gencode v12 (ref. 33) and miRBase v18 (ref. 34) annotations: protein-coding and lincRNA genes (16,084 detected in >50% of samples), transcripts (67,603; with FluxCapacitor⁷), exons (146,498), annotated splice junctions (129,805; analysed in detail in P.G.F. *et al.*, manuscript submitted), transcribed repetitive elements (47,409), and mature miRNAs (715). Data quality was assessed by sample correlations and read and gene count distributions, and technical variation was removed by PEER normalization³⁵ for the QTL and miRNA-mRNA correlation analyses¹¹. The samples clustered uniformly both before and after normalization. The genotype data was obtained from 1000 Genomes Project Phase 1 data set for 421 samples (80× average exome and 5× whole-genome read depth), and the remaining 41 samples were imputed from Omni 2.5M SNP array data. Furthermore, we did functional reannotation for all the 1000 Genomes Project variants using Gencode v12. QTL mapping was done with linear regression, using genetic variants with >5% frequency in 1-megabase window and normalized quantifications transformed to standard normal. Permutations were used to adjust the false discovery rate to 5%. Full details are provided in the Supplementary Methods.

individuals, and the vertical lines show the range of ratios for variants carried by several individuals. The grey vertical lines denote derived allele frequencies of 0, 0.001 and 0.01. **b**, Exon inclusion scores for variable exons for individuals that carry 0, 1 or 2 copies of variants that destroy a splice motif, with P-value from Mann-Whitney test.

Received 7 February; accepted 5 August 2013.

Published online 15 September 2013.

- Abecasis, G. R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
- The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
- Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
- Stranger, B. E. *et al.* Population genomics of human gene expression. *Nature Genet.* **39**, 1217–1224 (2007).
- Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nature Genet.* **44**, 1084–1089 (2012).
- Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773–777 (2010).
- Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010).
- Nica, A. C. *et al.* Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* **6**, e1000895 (2010).
- Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
- Hoehn, P. A. C. *et al.* Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. *Nature Biotech.* <http://dx.doi.org/10.1038/nbt.2702> (in the press).
- Gonzalez-Porta, M., Calvo, M., Sammeth, M. & Guigo, R. Estimation of alternative splicing variability in human populations. *Genome Res.* **22**, 528–538 (2012).
- Merklin, J., Russell, C., Chen, P. & Burge, C. B. Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science* **338**, 1593–1599 (2013).
- Barbosa-Morais, N. L. *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**, 1587–1593 (2012).
- Parts, L. *et al.* Extent, causes, and consequences of small RNA expression variation in human adipose tissue. *PLoS Genet.* **8**, e1002704 (2012).
- Xiao, C. & Rajewsky, K. MicroRNA control in the immune system: basic principles. *Cell* **136**, 26–36 (2009).
- Ebert, M. S. & Sharp, P. A. Roles for microRNAs in conferring robustness to biological processes. *Cell* **149**, 515–524 (2012).
- Pickrell, J. K., Pai, A. A., Gilad, Y. & Pritchard, J. K. Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet.* **6**, e1001236 (2010).
- Lee, Y. *et al.* Variants affecting exon skipping contribute to complex traits. *PLoS Genet.* **8**, e1002998 (2012).
- Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
- Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nature Rev. Genet.* **10**, 691–703 (2009).
- Veyrieras, J. B. *et al.* High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* **4**, e1000214 (2008).
- Gaffney, D. J. *et al.* Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol.* **13**, R7 (2012).
- McDaniell, R. *et al.* Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* **328**, 235–239 (2010).
- Degner, J. F. *et al.* DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**, 390–394 (2012).

26. Hindorf, L. A., Junkins, H. A., Hall, P. N., Mehta, J. P. & Manolio, T. A. A Catalog of Published Genome-Wide Association Studies; available at <http://www.genome.gov/gwastudies> (accessed 11 September 2012).
27. O'Seaghdha, C. M. *et al.* Common variants in the calcium-sensing receptor gene are associated with total serum calcium levels. *Hum. Mol. Genet.* **19**, 4296–4303 (2010).
28. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
29. Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.* **23**, 198–199 (1998).
30. Montgomery, S. B., Lappalainen, T., Gutierrez-Arcelus, M. & Dermitzakis, E. T. Rare and common regulatory variation in population-scale sequenced human genomes. *PLoS Genet.* **7**, e1002144 (2011).
31. Marco-Sola, S., Sammeth, M., Guigo, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nature Methods* **9**, 1185–1188 (2012).
32. Pantano, L., Estivill, X. & Marti, E. SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. *Nucleic Acids Res.* **38**, e34 (2010).
33. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
34. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152–D157 (2011).
35. Stegle, O., Parts, L., Durbin, R. & Winn, J. A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. *PLOS Comput. Biol.* **6**, e1000770 10.1371/journal.pcbi.1000770 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank E. Falconnet, L. Romano, A. Planchon, D. Bielsen, A. Yurovsky, A. Buil, J. Bryois, A. Nica, I. Topolsky, N. Fusi, S. Waszak, C. Bustamante, J. Rung, N. Kolesnikov, A. Roa, E. Bragin, S. Brent, J. Gonzalez, M. Morell, A. Puig, E. Palumbo, M. Ventayol Garcia, J. F. J. Laros, J. Blanc, R. Birkelund, G. Plaja, M. Ingham, J. Camps, M. Bayes, L. Agueda, A. Gouin, M.-L. Yaspo, E. Graf, A. Walther, C. Fischer, S. Loesecke, B. Schmick, D. Balzereit, S. Dökel, M. Linser, A. Kovacovics, M. Friskovec, C. von der Lancken, M. Schlapkohl, A. Hellmann, M. Schilhabel, the SNP&SEQ Technology Platform in Uppsala, S. Sauer, the Vital-IT high-performance computing centre of the SIB Swiss Institute of Bioinformatics, B. Goldstein and others at the Coriell Institute, and J. Cooper, E. Burnett, K. Ball and others at the European Collection of Cell Cultures (ECACC) and the 1000 Genomes Consortium. This project was funded by the European Commission 7th Framework Program (FP7) (261123; GEUVADIS); the Swiss National Science Foundation (130326, 130342), the Louis Jeantet Foundation, and ERC (260927) (E.T.D.); NIH-NIMH (MH090941) (E.T.D., M.I.M., R.G.); Spanish Plan Nacional SAF2008-00357 (NOVADIS), the Generalitat de Catalunya AGAUR 2009 SGR-1502, and the Instituto de Salud Carlos III (FIS/FEDER PI11/00733) (X.E.); Spanish Plan Nacional (BIO2011-26205) and ERC (294653) (R.G.); ESGL, READNA (FP7 Health-F4-2008-201418), Spanish Ministry of Economy

and Competitiveness (MINECO) and the Generalitat de Catalunya (I.G.G.); DFG Cluster of Excellence Inflammation at Interfaces, the INTERREG4A project HIT-ID, and the BMBF IHEC project DEEP SP 2.3 (P.Ro.); German Centre for Cardiovascular Research (DZHK) and the German Ministry of Education and Research (01GR0802, 01GM0867, 01GR0804, 16EX1020C) (T.M.); EurocanPlatform (FP7 260791), ENGAGE and CAGEKID (241669) (A.B.); FP7/2007-2013, ENGAGE project, HEALTH-F4-2007-201413, and the Centre for Medical Systems Biology within the framework of The Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific and Research (NWO) (P.A.C.H. and G.-J.v.O.); The Swedish Research Council (C0524801, A028001) and the Knut and Alice Wallenberg Foundation (2011.0073) (A.-C.S.); The Swiss National Science Foundation (127375, 144082) and ERC (249968) (S.E.A.); Instituto de Salud Carlos III (FIS/FEDER PS09/02368) (A.C.); German Federal Ministry of Education and Research (01GS08201) (R.S.); Max Planck Society (H.L.); Wellcome Trust (WT085532) and the European Molecular Biology Laboratory (P.F.); ENGAGE, Wellcome Trust (081917, 090367, 090532, 098381), and Medical Research Council UK (G0601261) (M.I.M.); Wellcome Trust Centre for Human Genetics (090532/Z/09/Z, 075491/Z/04/B), Wellcome Trust (098381, 090367, 076113, 083270), the WTCCC2 project (085475/B/08/Z, 085475/Z/08/Z), Royal Society Wolfson Merit Award, Wellcome Trust Senior Investigator Award (095552/Z/11/Z) (P.D.); EMBO long-term fellowship EMBO-ALTF 2010-337 (H.K.); NIH-NIGMS (R01 GM104371) (D.G.M.); Marie Curie FP7 fellowship (O.S.); Scholarship by the Clarendon Fund of the University of Oxford, and the Nuffield Department of Medicine (M.A.R.); EMBO long-term fellowship ALTF 225-2011 (M.R.F.); Emil Aaltonen Foundation and Academy of Finland fellowships (T.L.).

Author Contributions Designed the study: T.L., T.Gi., S.B.M., P.A.C.H., E.L., H.L., S.S., R.S., A.C., S.E.A., R.H., A.-C.S., G.-J.v.O., A.B., T.M., P.Ro., R.G., I.G.G., X.E. and E.T.D. Coordinated the project: T.L., T.Gi., G.B., X.E. and E.T.D. Participated in data production: T.L., T.Gi., I.Pa., M.Su., E.L., S.B., M.G., V.A., K.K., D.E., P.Ri. and O.K. Analysed the data: T.L., M.Sa., M.R.F., P.A.C.H., J.M., M.A.R., M.G.-P., N.K., T.Gr., P.G.F., M.B., T.W., L.G., M.v.I., J.A., P.Ri., I.Pu., D.E., A.T., M.Su., D.G.M., M.L., E.L., H.P.J.B., I.Pa., T.S., O.K., H.O., H.K., S.B., M.G., K.K., V.A., O.S., M.P., P.D., M.I.M., P.F. and T.M.S. Drafted the paper: T.L. and E.T.D. See Supplementary Note for Members of the Geuvadis Consortium.

Author Information The Geuvadis RNA-sequencing data, genotype data, variant annotations, splice scores, quantifications, and QTL results are freely and openly available with no restrictions. The main portal for accessing the data is EBI ArrayExpress, under accessions E-GEUV-1, E-GEUV-2 and E-GEUV-3 (see the data access schema in Supplementary Fig. 39). For visualization of the results we created the Geuvadis Data Browser (<http://www.ebi.ac.uk/Tools/geuvadis-das>) where quantifications and QTLs can be viewed, searched and downloaded (Supplementary Fig. 40). The project webpage (<http://www.geuvadis.org>) provides full documentation and links to all files, and the analysis group wiki is open to the public (<http://geuvadiswiki.crg.es>). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.L. (tuuli.e.lappalainen@gmail.com) or E.T.D. (emmanouil.dermizakis@unige.ch).

The ubiquitin ligase parkin mediates resistance to intracellular pathogens

Paolo S. Manzanillo¹, Janelle S. Ayres^{2*}, Robert O. Watson^{1*}, Angela C. Collins³, Gianne Souza¹, Chris S. Rae⁴, David S. Schneider⁵, Ken Nakamura^{6,7}, Michael U. Shiloh³ & Jeffery S. Cox¹

Ubiquitin-mediated targeting of intracellular bacteria to the autophagy pathway is a key innate defence mechanism against invading microbes, including the important human pathogen *Mycobacterium tuberculosis*. However, the ubiquitin ligases responsible for catalysing ubiquitin chains that surround intracellular bacteria are poorly understood. The parkin protein is a ubiquitin ligase with a well-established role in mitophagy, and mutations in the parkin gene (*PARK2*) lead to increased susceptibility to Parkinson's disease. Surprisingly, genetic polymorphisms in the *PARK2* regulatory region are also associated with increased susceptibility to intracellular bacterial pathogens in humans, including *Mycobacterium leprae* and *Salmonella enterica* serovar Typhi, but the function of parkin in immunity has remained unexplored. Here we show that parkin has a role in ubiquitin-mediated autophagy of *M. tuberculosis*. Both parkin-deficient mice and flies are sensitive to various intracellular bacterial infections, indicating parkin has a conserved role in metazoan innate defence. Moreover, our work reveals an unexpected functional link between mitophagy and infectious disease.

Eukaryotic cells target invading microbes to autophagosomes through a process termed xenophagy, which has a key role in innate immune defence. Various intracellular bacterial pathogens, including *Mycobacterium tuberculosis*, are targeted for xenophagy through a ubiquitin-mediated pathway that surrounds bacteria with conjugated ubiquitin chains^{1–3}. Marking with polyubiquitin presumably recruits ubiquitin-binding autophagy adaptors such as p62, which in turn engage the autophagic machinery for autophagosome formation and delivery of bacteria to the lysosome^{1,3,4}. Although ubiquitin-binding adaptors are required for xenophagy, whether ubiquitin itself directly mediates bacterial autophagy is not clear because the identities of the ubiquitinated substrate(s) and ligase(s) responsible for coating cytosol-exposed bacteria are poorly understood.

In a fashion similar to xenophagy, the process of mitophagy eliminates damaged mitochondria through ubiquitin-mediated targeting to autophagosomes. A key step in mitophagy is marking of damaged mitochondria by the ubiquitin ligase parkin, which localizes to the organelle and directly ubiquitinates proteins on the mitochondrial surface⁵. Ubiquitin-tagged mitochondria are directed to the autophagosome pathway by p62 (refs 6–8) and several other factors⁵, ultimately delivering the organelle to the lysosome⁵.

PARK2 mutations in humans are well-known risk factors for the development of Parkinson's disease, but polymorphisms in the regulatory region of *PARK2*, some of which result in decreased parkin expression⁹, have been associated with increased susceptibility to the intracellular pathogens *Mycobacterium leprae* and *S. enterica* serovar Typhi^{10,11}. Although a genetic link to increased infection risk has been identified, the function of parkin in immunity has remained obscure. We have identified that parkin, similar to its role in mitophagy, is also important for innate defence against *M. tuberculosis* and other intracellular pathogens by promoting xenophagy. This work provides a

possible mechanism underlying the human genetic studies linking parkin to increased susceptibility to bacterial infection and reveals a surprising connection between mitochondrial homeostasis and pathogen defence.

Parkin in TB-ubiquitin co-localization

We have shown previously that on infection of macrophages, *M. tuberculosis* bacilli that puncture phagosomal membranes through their ESX-1 secretion system gain access to the host cytosol but become enveloped by conjugated ubiquitin chains and are targeted to autophagosomes by p62 and NDP52 (ref. 3). Although the role of ESX-1 in autophagy induction is probably complicated¹², it is clear that approximately one-third of wild-type intracellular bacteria are targeted to autophagy during macrophage infection and that this has a major role in host resistance to infection^{2,3}. Because of the commonalities between mitophagy and autophagy of intracellular mycobacteria, and the links between *PARK2* polymorphisms and increased susceptibility to bacterial infection in humans, we speculated that parkin may also be recruited to *M. tuberculosis*-containing phagosomes and target them for ubiquitin-mediated autophagy. Indeed, after infection of murine bone marrow-derived macrophages (BMDMs) with *M. tuberculosis* expressing mCherry, we found that parkin localized to approximately 12% of wild-type *M. tuberculosis* phagosomes but not to ESX-1 mutants (Fig. 1a, Extended Data Fig. 1). Next, we infected BMDMs isolated from wild-type and *Park2*^{−/−} mice and performed immunofluorescence co-localization experiments using antibodies that recognize polyubiquitin. As shown in Fig. 1b, c, *Park2*^{−/−} BMDMs were severely defective for *M. tuberculosis* ubiquitin co-localization as compared to control macrophages, resulting in a significant reduction in ubiquitin-positive mycobacteria. Likewise, short hairpin RNA (shRNA) knockdown of parkin expression in human macrophage cell lines also resulted in a drastic reduction in ubiquitin

¹Department of Microbiology and Immunology Program in Microbial Pathogenesis and Host Defense, University of California, San Francisco, San Francisco, California 94158, USA. ²Immunobiology and Microbial Pathogenesis Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA. ³Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA. ⁴Department of Molecular and Cellular Biology, Division of Immunology and Pathogenesis, University of California, Berkeley, California 94720, USA. ⁵Department of Microbiology and Immunology, Stanford University, Stanford, California 94305, USA. ⁶Gladstone Institute of Neurological Disease, University of California, San Francisco, California 94158, USA. ⁷Department of Neurology and Graduate Programs in Neuroscience and Biomedical Sciences, University of California, San Francisco, California 94158, USA.

*These authors contributed equally to this work.

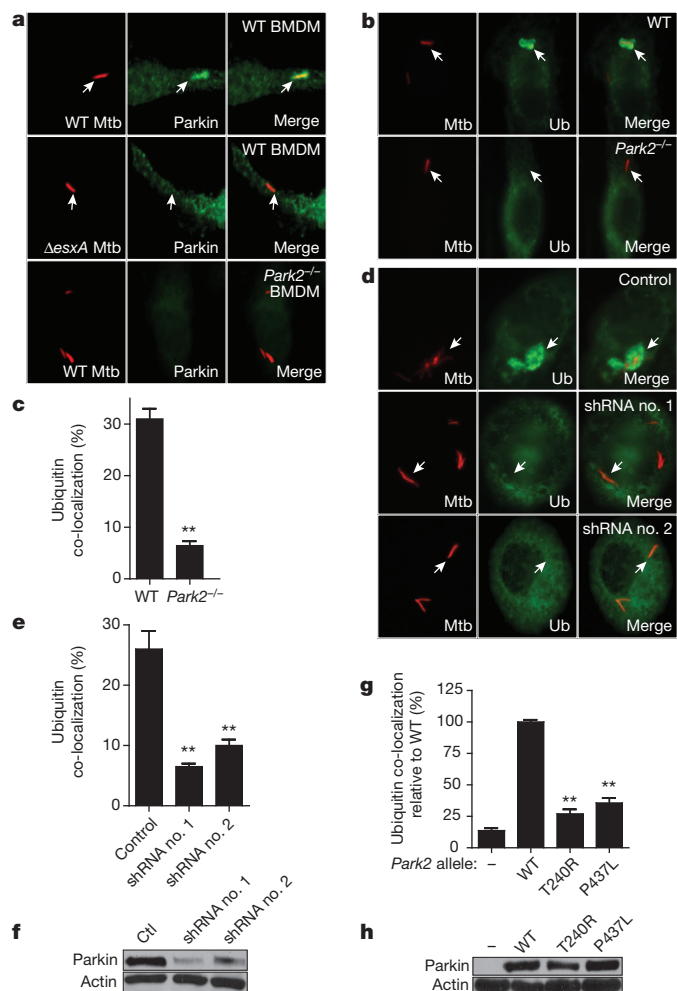


Figure 1 | Parkin activity is required for *M. tuberculosis*-ubiquitin co-localization. **a**, Wild-type (WT) and $Park2^{-/-}$ BMDMs were infected with mCherry-expressing *M. tuberculosis* (Mtb) for 4 h and immunostained using anti-parkin antibodies. **b**, Wild-type BMDMs were infected with mCherry-expressing *M. tuberculosis* for 4 h and immunostained for polyubiquitin. **c**, Quantification of ubiquitin-positive *M. tuberculosis* from **b**. Results are means \pm s.e.m. of three independent experiments (** $P < 0.001$, paired Student's *t*-test). **d**, U937 human macrophages expressing a scrambled shRNA (Control) or one of two different shRNAs targeting *PARK2* (shRNA no. 1, shRNA no. 2) were infected with mCherry-expressing *M. tuberculosis* for 12 h and immunostained for polyubiquitin. **e**, Quantification of ubiquitin-positive *M. tuberculosis* from **d**, results are means \pm s.e.m. of three independent experiments (** $P < 0.005$, Student's *t*-test). **f**, The parkin and actin expression in cells from **e** was determined by western blotting. **g**, $Park2^{-/-}$ BMDMs were transduced with lentivirus expressing BFP (-), wild-type parkin, or two separate mutant parkin alleles (T240R, P437L). Cells were infected with *M. tuberculosis* for 4 h and ubiquitin and *M. tuberculosis* co-localization was quantified and expressed relative to control BMDMs. Results are means \pm s.e.m. of three independent experiments (** $P < 0.005$, paired Student's *t*-test). **h**, The parkin and actin expression in cells from **g** was determined by western blotting.

localization with *M. tuberculosis* cells (Fig. 1d–f), indicating that parkin has a conserved role in mycobacterium ubiquitination in mice and humans. Knockdown of LRSAM1, a ubiquitin ligase recently implicated in antibacterial defence and ubiquitination of salmonellae^{1,3,4,13}, had no effect on ubiquitin or GFP-LC3 co-localization with *M. tuberculosis* (Extended Data Fig. 1b, c). Expression of wild-type *Park2* in $Park2^{-/-}$ cells restored ubiquitin localization around *M. tuberculosis* cells (Fig. 1g, h). In contrast, $Park2^{-/-}$ BMDMs expressing either of two pathogenic RING domain mutant alleles that inactivate parkin's E3 ligase activity, T240R or P437L^{3,4,14–16}, failed to restore ubiquitin co-localization with *M. tuberculosis* (Fig. 1g, h). Taken together, these data demonstrate

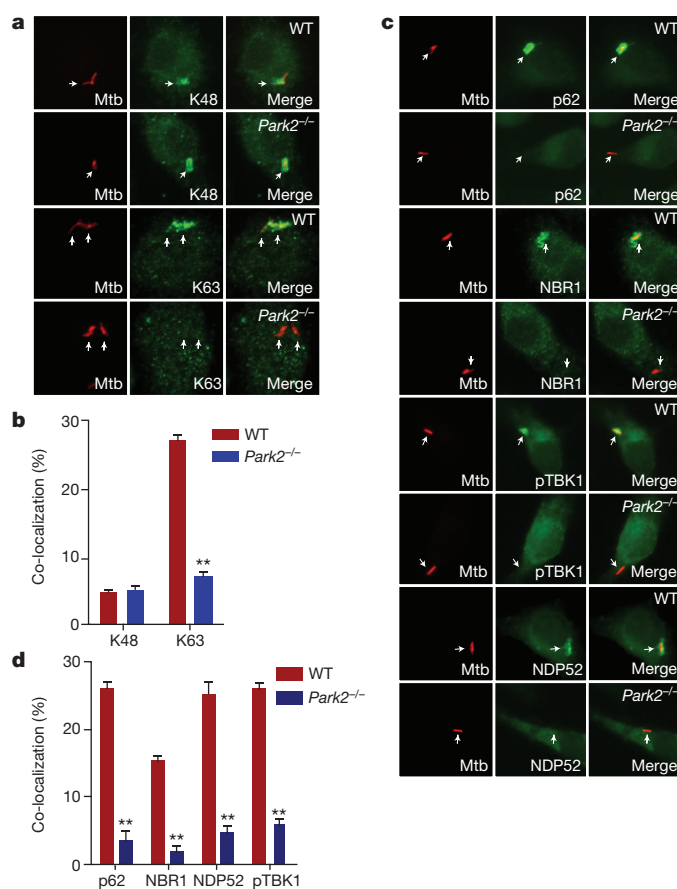


Figure 2 | Parkin mediates K63-ubiquitin co-localization of *M. tuberculosis* and recruitment of ubiquitin-autophagy receptors. **a**, BMDMs were infected with mCherry-expressing *M. tuberculosis* for 4 h and immunostained using linkage-specific ubiquitin antibodies. **b**, Quantification of K63 and K48 ubiquitin-positive *M. tuberculosis* from **a**. (** $P < 0.001$, paired Student's *t*-test). **c**, Same as **a** except immunostained for NDP52, p62, phospho-TBK1 or NBR1. **d**, Quantification of co-localization from **c** ($n = 3$ per group, ** $P < 0.001$). All errors, s.e.m.

that parkin and its E3 ligase activity are critical for the co-localization of ubiquitin with *M. tuberculosis* during infection.

Parkin mediates K63-linked polyubiquitin

We showed previously that both K63- and K48-linked polyubiquitin chains accumulate around *M. tuberculosis*³. Because parkin is known to catalyse K63-linked ubiquitin chains^{5,17}, we sought to determine the nature of the residual ubiquitin surrounding *M. tuberculosis* in $Park2^{-/-}$ BMDMs. Using ubiquitin linkage-specific antibodies^{5,18}, we found that in wild-type BMDMs, approximately 26–29% of all intracellular bacteria (~90–95% of all ubiquitin-positive bacilli) co-localized with K63 ubiquitin, whereas only 5–7% bacilli stained for K48 (Fig. 2a, b). Additionally, expression of HA-epitope-tagged forms of K48 and K63 ubiquitin within BMDMs supported the notion that K63-linked polyubiquitin is more abundant surrounding *M. tuberculosis* than the K48-linked form (Extended Data Fig. 2). In $Park2^{-/-}$ BMDMs, however, there was a specific decrease in the number of K63-positive mycobacteria, whereas the K48-positive population remained unaffected (Fig. 2a, b, Extended Data Fig. 2). Previous electron microscopy studies indicated that though ubiquitin can localize directly with *M. tuberculosis*, the majority of ubiquitin is found on membranous structures surrounding *M. tuberculosis*-containing phagosomes^{15,19–21}. To address whether bacterial or host proteins become ubiquitinated, we used a digitonin-based method that specifically permeabilizes the plasma membrane and leaves phagosomes intact

with luminal contents inaccessible to antibodies^{10,22} (Extended Data Fig. 3a). As shown in Extended Data Fig. 3b, antibodies against poly-ubiquitin and K63-ubiquitin stained digitonin-permeabilized cells and resulted in co-localization with mCherry-expressing *M. tuberculosis*. Importantly, anti-*M. tuberculosis* antibodies failed to stain *M. tuberculosis* within digitonin-permeabilized cells and stained cells only after addition of Triton X-100 detergent, demonstrating that digitonin-permeabilized cells contained intact phagosomes (Extended Data Fig. 3b, c). Taken together, these data suggest that parkin facilitates the linkage of K63-linked ubiquitin chains surrounding *M. tuberculosis* containing phagosomes, although the exact protein target(s) remain to be explored. Furthermore, these data also suggest that at least one other ubiquitin ligase works independently of parkin to catalyse the K48-linked ubiquitination that surrounds a minor population of *M. tuberculosis* cells.

Parkin required for TB autophagy

Ubiquitination coincides with autophagic targeting of *M. tuberculosis*, but a causal relationship has not been demonstrated. To determine whether parkin-mediated ubiquitination directs autophagic targeting of *M. tuberculosis*, we infected wild-type and *Parkin*^{-/-} macrophages with *M. tuberculosis* and measured co-localization of bacilli with multiple markers of autophagy. Microscopy analysis of proteins involved in ubiquitin recognition (NBR1, NDP52, p62, phospho-TBK1) revealed reduced co-localization with *M. tuberculosis* in *Parkin*^{-/-} macrophages (Fig. 2c, d), suggesting that parkin-mediated ubiquitination directly leads to the recruitment of the proximal ubiquitin-adaptors that facilitate autophagic targeting of mycobacteria. Likewise, mycobacterial cells within infected *Parkin*^{-/-} BMDMs had reduced co-localization with autophagic proteins LC3 and ATG12 relative to infection of wild-type BMDMs (Fig. 3a, b), suggesting that the K63-linked poly-ubiquitin catalysed by parkin is required for delivery of *M. tuberculosis* to autophagosomes. Consistent with this notion, *Parkin*^{-/-} cells were defective in conversion of LC3 to its activated, lipidated form, LC3-II, during *M. tuberculosis* infection, further demonstrating that parkin is required for autophagy of mycobacteria (Fig. 3c).

Parkin limits TB replication

The autophagy pathway serves to limit *M. tuberculosis* replication in macrophages by delivering bacilli to the lysosome^{3,11}. To determine whether parkin-mediated ubiquitination is required for autophagic targeting of *M. tuberculosis* to lysosomes, we infected BMDMs with *M. tuberculosis* and monitored co-localization with the lysosomal marker, LAMP1. During *M. tuberculosis* infection of wild-type BMDMs, approximately 30% of bacilli stained positive for LAMP1 at 6 h post-infection (Fig. 3d, e). In contrast, only 2–5% of bacilli co-localized with LAMP1 during *M. tuberculosis* infection of *Parkin*^{-/-} macrophages. This was similar to macrophages deficient for the essential autophagy protein, ATG5 (Fig. 3d, e)^{3,5}. To test whether these differences led to changes in bacterial survival, we infected *Parkin*^{-/-} and *Atg5*^{-/-} BMDMs with wild-type *M. tuberculosis* and determined bacterial viability by enumerating colony-forming units (c.f.u.). Infection of BMDMs deficient for either ATG5 or parkin resulted in a 2- or a 2.5-fold increase in bacterial numbers, respectively, relative to control BMDMs by 12 h post-infection (Fig. 3f). Conversely, overexpression of parkin in RAW 264.7 macrophages led to decreased bacterial replication (Fig. 3g). Importantly, knockdown of parkin expression in human U937 cells also led to an increase in bacterial replication during infection (Fig. 3h). Taken together, our data demonstrate that parkin-mediated ubiquitination leads to the autophagic targeting of *M. tuberculosis* and is essential for inhibition of mycobacterial replication in macrophages.

Parkin mediates *M. tuberculosis* immunity

Polymorphisms within the regulatory region of *PARK2* in human populations have been identified as a common risk factor for increased susceptibility to *Mycobacterium leprae* and salmonellae infection^{6,7,10,11}, suggesting that parkin has an important role *in vivo* against a broad

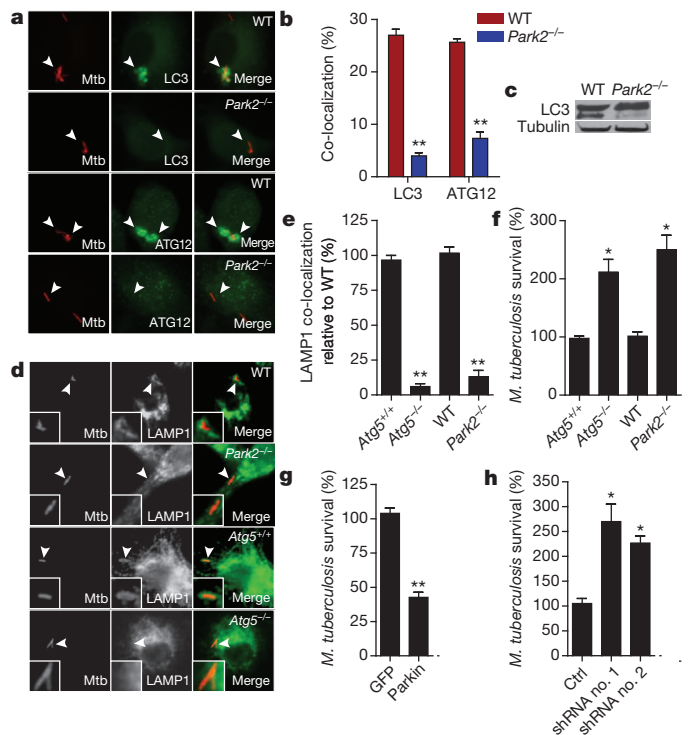


Figure 3 | Parkin mediates autophagic targeting of *M. tuberculosis* and limits bacterial replication. **a**, BMDMs were infected with mCherry-expressing *M. tuberculosis* (arrows) for 4 h and immunostained for LC3 (LC3B antibody) or ATG12. **b**, Quantification of results from **a** ($n = 3$ per group, $^{***}P < 0.001$). **c**, Western blot analysis of LC3 (LC3B antibody) from cell lysates from **a**. **d**, BMDMs were infected with mCherry-expressing *M. tuberculosis* (arrows) for 6 h and immunostained for LAMP1. **e**, Quantification of results from **d** expressed relative to control BMDMs ($n = 3$ per group, $^{***}P < 0.001$). **f**, BMDMs were infected with *M. tuberculosis* and colony-forming units (c.f.u.) at $t = 0$ and $t = 16$ h were determined by plating. Results were normalized to $t = 0$ ($n = 3$ per group, $^{*}P < 0.02$). **g**, RAW264.7 macrophages transduced with lentivirus expressing green fluorescent protein (GFP) or parkin were infected with *M. tuberculosis* and c.f.u. were determined after 16 h. Results were normalized to GFP-expressing cells ($n = 3$ per group, $^{***}P < 0.0076$). **h**, U937 human macrophages expressing either scrambled or *PARK2* shRNAs were infected with *M. tuberculosis* for 36 h and c.f.u. were determined. Results were normalized to $t = 0$ ($N = 3$ per group, $^{*}P < 0.02$). All errors, s.e.m.

range of intracellular bacterial infections. We began to test this by first determining whether parkin was required *in vivo* during *M. tuberculosis* infection of mice. We performed a low-dose aerosol infection of wild-type and *Parkin*^{-/-} knockout mice and determined mouse survival and bacterial burden within infected tissues. In comparison to infected wild-type mice, *Parkin*^{-/-} knockout mice had a tenfold increase in bacterial c.f.u. within infected lungs, spleens and liver by 21 days post-infection (Fig. 4a, b). Furthermore, survival studies revealed that *Parkin*^{-/-} mice were extremely susceptible to *M. tuberculosis*, because all infected mice succumbed to overwhelming infection by 85 days post-infection, whereas all infected wild-type mice remained alive and displayed no overt signs of weight loss or stress (Fig. 4c). Immunohistochemistry staining of infected mouse lungs revealed robust parkin expression in mouse granulomas within the central macrophage-containing zone (Fig. 4d). In agreement with our mouse experiments, we also observed high expression of parkin within human lung granuloma tissue samples from *M. tuberculosis*-infected patients (Fig. 4e, Extended Data Fig. 4). Further analysis of human lung specimens by confocal microscopy revealed the presence of parkin puncta within *M. tuberculosis*-infected cells as well as *in vivo* co-localization of parkin with *M. tuberculosis* (Fig. 4f). *Parkin*^{-/-} mice were also highly susceptible to another intracellular pathogen, *Listeria monocytogenes*, resulting in 10–20-fold higher bacterial burdens relative to wild-type

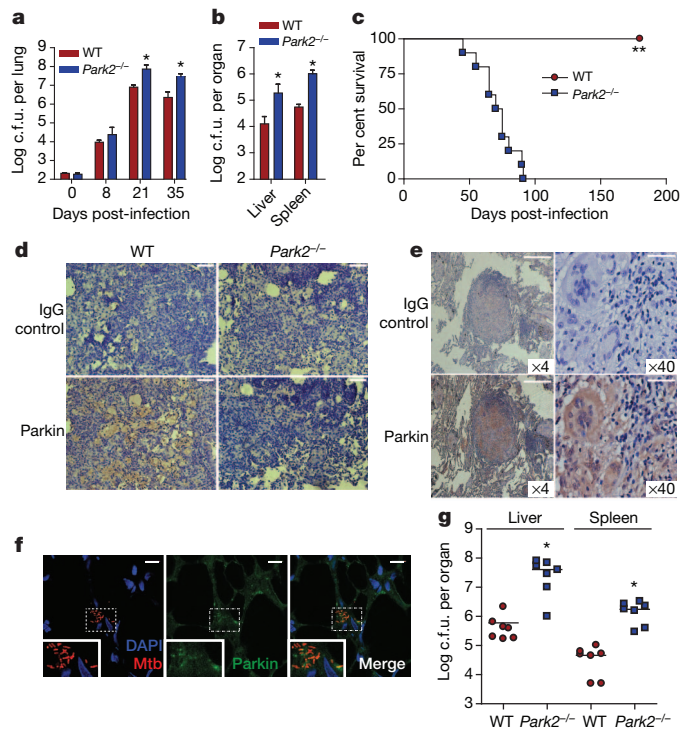


Figure 4 | Parkin is required for control of bacterial infection *in vivo*. **a**, WT and *Park2*^{-/-} mice were infected with wild-type *M. tuberculosis* by aerosol and lung bacterial burdens were determined by plating (means \pm s.d., $n = 5$ per group, $*P < 0.02$ by Student's *t*-test). **b**, Enumeration of liver and spleen c.f.u. from mice infected in **a**, 21 days post-infection (means \pm s.d., $n = 5$ per group, $*P < 0.03$ by Student's *t*-test). **c**, Survival of *M. tuberculosis*-infected wild-type and *Park2*^{-/-} mice ($n = 10$, $**P < 0.001$ by log-rank test). **d**, Immunohistochemistry staining of lung sections from infected mice 21 days post-infection, scale bar, 50 μ m. **e**, Immunohistochemistry staining of lung sections from a human patient with active tuberculosis. Scale bars, 500 μ m ($\times 4$), 50 μ m ($\times 40$). **f**, Confocal microscopy images of human lung sections from **e** immunostained for *M. tuberculosis* and parkin; DNA was visualized using DAPI (4',6-diamidino-2-phenylindole). Scale bars, 25 μ m. **g**, Mice were infected with wild-type *L. monocytogenes* by i.p. injection and bacterial burdens in livers and spleens were determined by plating (means \pm s.d., $n = 7$ per group, $*P < 0.04$ by Student's *t*-test).

mice within infected spleens and liver (Fig. 4g). Taken together, these data demonstrate that parkin is essential *in vivo* for controlling intracellular bacterial pathogens within mice and suggest a role for parkin in human tuberculosis disease.

Conserved role of parkin in immunity

Parkin is present in all metazoans^{8,23}, including *Drosophila melanogaster* and *Caenorhabditis elegans*, with well-characterized functions in mitochondrial maintenance and in models of Parkinson's disease. Because xenophagy of intracellular pathogens is a highly evolutionarily conserved innate immune defence mechanism²⁴, we sought to determine whether parkin also has an evolutionarily conserved role in immunity within non-mammalian organisms. We began by first analysing *parkin*-deficient *D. melanogaster* strains using models of bacterial systemic infection. We obtained two mutant fly lines with independent disruptions of the *parkin* gene (*park*⁰¹⁹⁵⁰, *park*⁰⁰⁰⁶²) and infected them with *L. monocytogenes*, which has previously been shown to induce autophagy within flies^{25,26}. In contrast to wild-type infected flies, *parkin* mutants were severely defective in ATG8/LC3 processing during infection (Fig. 5a), suggesting that *parkin* has a role in autophagic immunity within flies. Consistent with our results in mice, *parkin* mutant flies were also highly susceptible to *L. monocytogenes* infection and led to 10–50-fold increases in bacterial burdens relative to wild-type infected flies (Fig. 5b). This was accompanied by

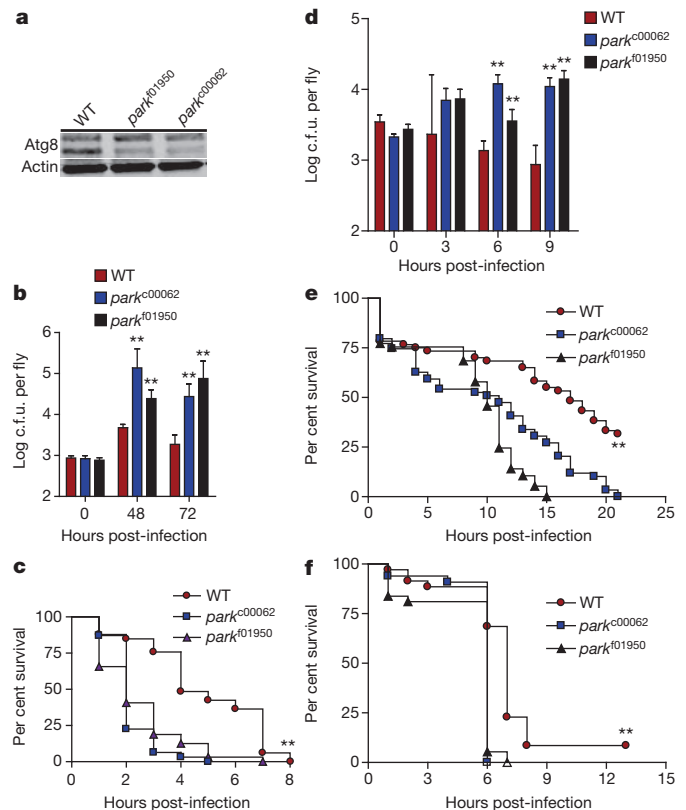


Figure 5 | Parkin is required for control of *S. enterica* serovar Typhimurium and *M. marinum* infection within flies. **a**, Wild-type and two independent *parkin*-deficient *D. melanogaster* lines (*park*⁰⁰⁰⁶², *park*⁰¹⁹⁵⁰) were infected with *L. monocytogenes* by anterior abdomen injection. Atg8 processing was monitored in whole-fly protein lysates by western blotting. **b**, *L. monocytogenes* c.f.u. from **a** were determined by plating (means \pm s.d., $n = 3$ –5, $**P < 0.001$ by Student's *t*-test). **c**, Survival of *D. melanogaster* was determined ($n = 33$, $**P < 0.001$ by log-rank test). **d**, **e**, Flies were infected with *S. enterica* serovar Typhimurium and bacterial c.f.u. (**d**) (means \pm s.d., $n = 3$ –5 per group, $**P < 0.009$ by Student's *t*-test) and fly survival (**e**) were determined ($n = 59$, $**P < 0.001$ by log-rank test). **f**, Survival of flies infected with *M. marinum* ($n = 37$, $**P < 0.0045$ by log-rank test).

decreased survival, with a median lifespan of two days following infection (Fig. 5c). In addition, *parkin* mutant flies were also susceptible to other autophagy-inducing intracellular pathogens such as *S. enterica* serovar Typhimurium and *Mycobacterium marinum*. By 9 h post-infection, *parkin*-deficient flies had a tenfold increase in *S. enterica* serovar Typhimurium burden relative to wild-type flies (Fig. 5d). Moreover, *parkin* mutant flies had significantly decreased lifespans upon infection with either *S. enterica* serovar Typhimurium (Fig. 5e) or *M. marinum* (Fig. 5f). *C. elegans* strains deficient in the *parkin* homologue (*pdr-1*) were also highly susceptible to *S. enterica* serovar Typhimurium infection (data not shown). Altogether, our data show that *parkin* homologues within metazoans are required for proper immune response to infection, suggesting an evolutionarily conserved role in innate immunity.

Discussion

Our findings reveal that parkin regulates a common cellular program by which metazoans mediate quality control of endogenous mitochondria (self) and eradicate harmful bacterial pathogens (non-self). Although these two activities are seemingly disparate, the evolutionary origin of mitochondria from a bacterial endosymbiont suggests that perhaps mitochondrial dysfunction triggers the recognition of the organelle as non-self. For example, mitochondria (and bacterial endosymbionts^{5,27}) may actively evade parkin surveillance, but these inhibitory processes are overridden on organelle damage. Alternatively, several studies have shown

that damaged mitochondria may serve as a 'danger-associated molecular pattern', resulting in the activation of several innate immune receptors such as Toll-like receptors and the NLRP3-inflammasome complex^{28,29}. In view of recent studies showing that intracellular infection with several pathogens such as *L. monocytogenes* result in altered host mitochondria dynamics³⁰, it is tempting to speculate that damaged mitochondria may serve as a signal for intracellular infection and activation of xenophagy.

Our results also provide a molecular explanation for increased bacterial susceptibility of humans with polymorphisms in the *PARK2* regulatory region^{3,10,11}, broadening the role of parkin beyond mitochondrial homeostasis. Indeed, several recent studies have shown parkin to participate in a wide array of cellular processes such as apoptosis³¹, regulation of lipid metabolism³² and cytokine production on infection³³, all of which may contribute to the *in vivo* importance of parkin in immunity. This work highlights the unexpected connection between mitochondrial-based neuronal disorders and susceptibility to bacterial infection in humans. Recent genome-wide association studies on inflammatory bowel disease, which is linked to altered host-gut microbe interactions, have identified susceptibility single nucleotide polymorphisms within *LRRK2* and *PARK7*, two genes canonically associated with Parkinson's disease^{34,35}. Thus, we surmise that genes typically associated with neuronal maintenance or mitophagy may have broad roles in cellular homeostasis within various cell types.

METHODS SUMMARY

Macrophages were infected with *M. tuberculosis* (Erdman strain) using a 'spin-infection' protocol³ in which monolayers were overlaid with bacterial suspensions, centrifuged for 10 min at 183g, and extensively washed to remove extracellular bacteria. *Park2*^{-/-} macrophages expressing parkin constructs were generated by transducing marrow cells from knockout mice with lentivirus, followed by differentiation into macrophages. During day 3 of differentiation, cells were selected with 5 µg ml⁻¹ of puromycin. For immunofluorescence, macrophages were cultured on coverslips, infected at a multiplicity of infection (m.o.i.) of 1 and fixed in 4% paraformaldehyde before permeabilization and incubation with antibodies. All co-localization studies were blinded, with a minimum count of 200 cells per coverslip, and were performed in triplicate. For c.f.u. experiments, macrophages were infected at m.o.i. = 5 and viable bacteria were enumerated by plating lysates on solid media. Mice (C57BL/6 background) were infected with *M. tuberculosis* by aerosol infection (200 c.f.u.) and animals were euthanized on losing 15% of their maximal body weight, in accordance with an approved UCSF IACUC protocol. For *L. monocytogenes* (10403 s strain) infections, mice were infected by intraperitoneal injection with 4 × 10⁵ c.f.u. Bacteria were enumerated by plating organ homogenates onto solid media. Male 5- to 7-day-old flies were injected with *L. monocytogenes* (1,000 c.f.u.), *S. enterica* serovar Typhimurium (2,500 c.f.u., SL1344 strain) or *M. marinum* (1,000 c.f.u., M strain) in 50 nl of culture into the anterior abdomen. Fly survival was calculated daily.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 November 2012; accepted 15 August 2013.

Published online 4 September 2013.

- Zhao, Z. *et al.* Autophagosome-independent essential function for the autophagy protein Atg5 in cellular immunity to intracellular pathogens. *Cell Host Microbe* **4**, 458–469 (2008).
- Deretic, V. & Levine, B. Autophagy, immunity, and microbial adaptations. *Cell Host Microbe* **5**, 527–549 (2009).
- Watson, R. O., Manzanillo, P. S. & Cox, J. S. Extracellular *M. tuberculosis* DNA targets bacteria for autophagy by activating the host DNA-sensing pathway. *Cell* **150**, 803–815 (2012).
- Wild, P. *et al.* Phosphorylation of the autophagy receptor optineurin restricts *Salmonella* growth. *Science* **333**, 228–233 (2011).
- Youle, R. J. & Narendra, D. P. Mechanisms of mitophagy. *Nature Rev. Mol. Cell Biol.* **12**, 9–14 (2011).
- Komatsu, M. *et al.* Homeostatic levels of p62 control cytoplasmic inclusion body formation in autophagy-deficient mice. *Cell* **131**, 1149–1163 (2007).
- Geisler, S. *et al.* PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. *Nature Cell Biol.* **12**, 119–131 (2010).
- Pankiv, S. *et al.* p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. *J. Biol. Chem.* **282**, 24131–24145 (2007).

- Chopra, R. *et al.* Mapping of PARK2 and PACRG overlapping regulatory region reveals LD structure and functional variants in association with leprosy in unrelated Indian population groups. *PLoS Genet.* **9**, e1003578 (2013).
- Mira, M. T. *et al.* Susceptibility to leprosy is associated with *PARK2* and *PACRG*. *Nature* **427**, 636–640 (2004).
- Ali, S. *et al.* *PARK2/PACRG* polymorphisms and susceptibility to typhoid and paratyphoid fever. *Clin. Exp. Immunol.* **144**, 425–431 (2006).
- Romagnoli, A. *et al.* ESX-1 dependent impairment of autophagic flux by *Mycobacterium tuberculosis* in human dendritic cells. *Autophagy* **8**, 1357–1370 (2012).
- Huett, A. *et al.* The LRR and RING domain protein LRSAM1 is an E3 ligase crucial for ubiquitin-dependent autophagy of intracellular *Salmonella* Typhimurium. *Cell Host Microbe* **12**, 778–790 (2012).
- Ponpuak, M. *et al.* Delivery of cytosolic components by autophagic adaptor protein p62 endows autophagosomes with unique antimicrobial properties. *Immunity* **32**, 329–341 (2010).
- Martin, I., Dawson, V. L. & Dawson, T. M. Recent advances in the genetics of Parkinson's disease. *Annu. Rev. Genomics Hum. Genet.* **12**, 301–325 (2011).
- Chen, D. *et al.* Parkin mono-ubiquitinates Bcl-2 and regulates autophagy. *J. Biol. Chem.* **285**, 38214–38223 (2010).
- Lim, K.-L. *et al.* Parkin mediates nonclassical, proteasomal-independent ubiquitination of synphilin-1: implications for Lewy body formation. *J. Neurosci.* **25**, 2002–2009 (2005).
- Newton, K. *et al.* Ubiquitin chain editing revealed by polyubiquitin linkage-specific antibodies. *Cell* **134**, 668–678 (2008).
- Kitada, T. *et al.* Mutations in the *parkin* gene cause autosomal recessive juvenile parkinsonism. *Nature* **392**, 605–608 (1998).
- Houben, D. *et al.* ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria. *Cell. Microbiol.* **14**, 1287–1298 (2012).
- Alonso, S., Pethe, K., Russell, D. G. & Purdy, G. E. Lysosomal killing of *Mycobacterium* mediated by ubiquitin-derived peptides is enhanced by autophagy. *Proc. Natl Acad. Sci. USA* **104**, 6031–6036 (2007).
- Collins, C. A. *et al.* Atg5-independent sequestration of ubiquitinated mycobacteria. *PLoS Pathog.* **5**, e1000430 (2009).
- Marin, I. & Ferrús, A. Comparative genomics of the RBR family, including the Parkinson's disease-related gene *parkin* and the genes of the *ariadne* subfamily. *Mol. Biol. Evol.* **19**, 2039–2050 (2002).
- Moy, R. H. & Cherry, S. Antimicrobial autophagy: a conserved innate immune response in *Drosophila*. *J. Innate Immun.* **5**, 444–455 (2013).
- Narendra, D., Kane, L. A., Hauser, D. N., Fearnley, I. M. & Youle, R. J. p62/SQSTM1 is required for parkin-induced mitochondrial clustering but not mitophagy; VDAC1 is dispensable for both. *Autophagy* **6**, 1090–1106 (2010).
- Yano, T. *et al.* Autophagic control of listeria through intracellular innate immune recognition in *Drosophila*. *Nature Immunol.* **9**, 908–916 (2008).
- Voronin, D., Cook, D. A. N., Steven, A. & Taylor, M. J. Autophagy regulates Wolbachia populations across diverse symbiotic associations. *Proc. Natl Acad. Sci. USA* **109**, E1638–E1646 (2012).
- Zhou, R., Yazdi, A. S., Menu, P. & Tschopp, J. A role for mitochondria in NLRP3 inflammasome activation. *Nature* **469**, 221–225 (2011).
- Nakahira, K. *et al.* Autophagy proteins regulate innate immune responses by inhibiting the release of mitochondrial DNA mediated by the NALP3 inflammasome. *Nature Immunol.* **12**, 222–230 (2011).
- Stavru, F., Bouillaud, F., Sartori, A., Ricquier, D. & Cossart, P. *Listeria monocytogenes* transiently alters mitochondrial dynamics during infection. *Proc. Natl Acad. Sci. USA* **108**, 3612–3617 (2011).
- Johnson, B. N., Berger, A. K., Cortese, G. P. & LaVoie, M. J. The ubiquitin E3 ligase parkin regulates the proapoptotic function of Bax. *Proc. Natl Acad. Sci. USA* **109**, 6283–6288 (2012).
- Kim, K.-Y. *et al.* Parkin is a lipid-responsive regulator of fat uptake in mice and mutant human cells. *J. Clin. Invest.* **121**, 3701–3712 (2011).
- de Léséleuc, L. *et al.* *PARK2* mediates interleukin 6 and monocyte chemoattractant protein 1 production by human macrophages. *PLoS Negl. Trop. Dis.* **7**, e2015 (2013).
- Anderson, C. A. *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nature Genet.* **43**, 246–252 (2011).
- Jostins, L. *et al.* Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).

Acknowledgements We thank N. Mizushima, S. Cherry, and K. Huynh for mice and reagents. We are grateful to S. Johnson for use of his microscope, members of the Schneider laboratory for assistance with fly work and D. Portnoy, R. Vance and S. Virgin for helpful discussions. This work was supported by National Institutes of Health grants R01 AI081727, P01 AI063302 and R01 AI099439, and NINDS P30NS069496 to K.N.

Author Contributions A.C.C. and M.U.S. performed immunohistochemistry staining of tissues and confocal microscopy of human lungs. P.S.M., C.S.R. and G.S. performed *Listeria* infections. J.S.A. performed all experiments involving *Drosophila melanogaster*. R.O.W. performed fluorescence microscopy experiments. P.S.M. performed all experiments involving *M. tuberculosis*. K.N. and D.S.S. provided reagents and resources. P.S.M. and J.S.C. conceived the study, designed the experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.S.C. (jeffery.cox@ucsf.edu).

METHODS

Mice and macrophages. *Park2*^{-/-} mice on the C57L/B6 background were a gift from K. Nakamura (Gladstone Institute)^{12,36}. GFP-LC3 transgenic mice on the C57L/B6 background were a gift from N. Mizushima and used as previously described³. Wild-type C57BL/6 mice were purchased from Jackson Laboratories. BMDMs were obtained from mouse femurs as previously described³ and cultured in DMEM H-21 supplemented with 20% FBS and 10% MCSF derived from 3T3-MCSF cells. U937 and Raw264.7 cells were obtained from ATCC and tested for mycoplasma before purchase. U937 monocytes were stimulated overnight with 20 ng ml⁻¹ of phorbol 12-myristate 13-acetate (PMA) (Sigma) before infections.

Antibodies. The following antibodies were used: LC3B (L10382 Invitrogen), NDP52 (Abnova D01), p62/SQSTM1 (Abnova 2C11), rabbit anti-*Drosophila* Atg-8 (gift from S. Cherry, University of Pennsylvania), anti-Mtb (BEI Resources NR-13818), polyubiquitin (Enzo FK1), phospho-TBK1 (Cell Signaling #5483), tubulin (Cell Signaling #2128), mouse-monoclonal parkin (Cell Signaling #4211) and ATG12 (Cell Signaling #2011), rabbit polyclonal anti-parkin (Abcam 15954), anti-NBR1 (Abcam 55474), humanized rabbit monoclonal antibodies specific for K63 or K48 (gift from E. Brown laboratory at Genentech).

Bacterial strains. The following bacterial strains were used: *M. tuberculosis* (Erdman), *L. monocytogenes* (10403 s), *M. marinum* (M), and *S. enterica* serovar Typhimurium (SL1344). Wild-type and Δ esxA mycobacteria expressing mCherry were previously described³.

Macrophage infection. For infections with *M. tuberculosis*, macrophages were infected as previously described³. Briefly, *M. tuberculosis* cultures were washed twice with PBS, gently sonicated to disperse clumps and resuspended in DMEM supplemented with 10% horse serum. Media was removed from cells, monolayers overlaid with the bacterial suspension and centrifuged for 10 min at 183g. Cells were washed twice in PBS and returned to macrophage media. For determination of bacterial viability following infection, cells were lysed in 1% Triton X-100 and plated on 7H10 solid media.

Western blotting. Protein lysates from cells and flies were obtained by lysis in RIPA buffer (Sigma) at the indicated time points. A micro BCA protein kit (Pierce) was used to measure protein levels and equal amounts of protein were electrophoresed on 4–20% Tris-HCL Criterion gels (Bio-Rad) and transferred onto nitrocellulose membranes. Western blots were analysed using an Odyssey Imager (Licor) according to the manufacturer's instructions. Western blot figures are a representative of at least two independent experiments.

Immunofluorescence. Infected cells were immunostained and visualized as previously described³. Briefly, macrophages were seeded onto poly-lysine-coated coverslips and infected with *M. tuberculosis* as described above. Cells were infected at an m.o.i. of 1, and fixed in 4% paraformaldehyde for 20 min at the indicated time points. Cells were incubated with indicated primary antibodies for 2 h at room temperature in 5% milk and 0.05% saponin, and were visualized using secondary Alexa-Fluor488 antibodies. For parkin immunofluorescence, cells were stained using rabbit polyclonal anti-parkin (Abcam15954) and an HRP conjugated donkey anti-rabbit secondary (1:100, Jackson Immunochemicals) followed by amplification with AlexaFluor-488 tyramide (Invitrogen). Co-localization studies were performed as blinded experiments, with a minimum count of 200 cells per coverslip and performed in triplicate. Data shown are the mean \pm s.e.m. of at least 3 experiments.

Mouse infection. Non-randomized mice (female 5–8-weeks old) were infected with *M. tuberculosis* through low-dose aerosol infection (200 c.f.u.) as previously described³⁷. Sample size ($n = 5$ per group) was based on empirical evidence from previously published reports³⁷ as the size necessary for adequate statistical analysis. Lungs, liver and spleens were collected, homogenized and plated on 7H10 agar plates. For survival experiments, infected mice were euthanized when they had lost 15% of their maximal body weight. For *L. monocytogenes* infections, mice were infected by intraperitoneal injection with 4×10^5 bacteria. Then 96 h post-infection, liver and spleen from infected mice were homogenized and serial dilutions were plated onto BHI agar plates. No blinding was done for these animal studies and all mice were housed and treated humanely using procedures described in animal care protocols approved by the Institutional Animal Care and Use Committees of UC San Francisco and UC Berkeley.

Fly strains and infections. The white118 strain (Bloomington stock centre, stock 6326) was used as the wild-type parental strain for all experiments. The *park*^{C00062} and *park*⁰¹⁹⁵⁰ alleles are from the Exelixis piggyBac transposon collection³⁸. *park*^{C00062} was obtained from Bloomington stock centre and *park*⁰¹⁹⁵⁰ was obtained from the Exelixis collection at Harvard. Infections were done as previously described³⁹. Male 5–7-day-old flies were anaesthetized with CO₂ and injected with

L. monocytogenes (1,000 c.f.u.), *S. enterica* serovar Typhimurium (2,500 c.f.u.) or *M. marinum* (1,000 c.f.u.) in 50 nl of culture into the anterior abdomen. Infected flies were homogenized in PBS supplemented with 1% Triton X-100 and serial dilutions were plated onto solid media. For survival analysis, the number of dead flies was counted daily and analysed using a log-rank test.

Lentiviral virus knockdown and complementation. Lentivirus expressing shRNAs targeting human *PARK2* and *LRSAM1* transcripts was generated using the Mission PLKO.1 lentivirus system from Sigma. A lentivirus expressing a non-targeting scrambled shRNA was used as a control. U937 cells were transduced with lentivirus as per the manufacturer's instructions and stable cell lines were generated by selection with puromycin. For transgene expression of human *PARK2* and *PARK2* mutants, full-length *PARK2* was cloned into pBluescript vector and RING domain mutants were generated using a QuikChange site-directed mutagenesis kit (Stratagene). HA-tagged ubiquitin constructs were obtained from Addgene³¹. Lentivirus expressing protein constructs was generated using the pLentiX system³. RAW264.7 cells were transduced and selected on puromycin for 1 week. BMDM cells expressing lentiviral constructs were generated by transducing marrow cells from knockout mice with lentivirus followed by differentiation into macrophages as described above. During day 3 of differentiation, cells were selected with 5 μ g ml⁻¹ of puromycin.

Statistics. Statistical analysis of data was performed using Prism software (GraphPad). Two-tailed unpaired Student's *t*-tests were used for analysis of microscopy images and mycobacterium c.f.u. assays. The Kaplan–Meir method was used to analyse mouse survival.

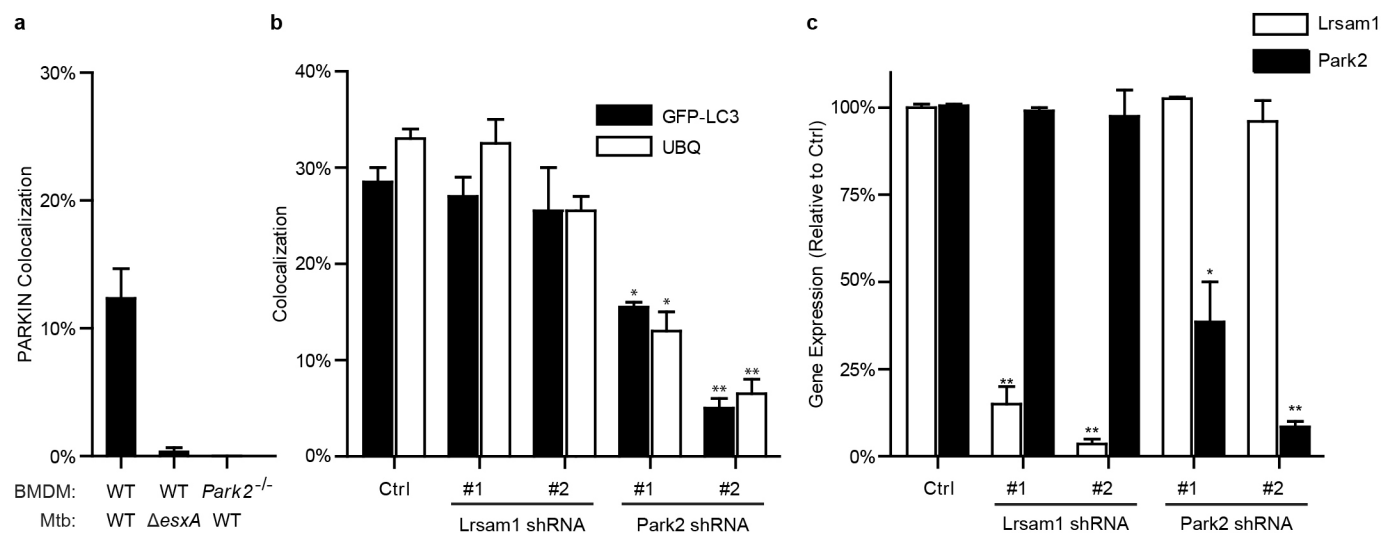
RNA isolation and qPCR. RNA was isolated and purified from cells using the Trizol micro-midi RNA isolation kit (Invitrogen) as per the manufacturer's instructions. For qPCR analysis, 2 μ g of RNA was reverse transcribed using the VILO cDNA synthesis kit (Invitrogen) and qPCR analysis was performed in triplicate using a MJ Research Opticon 3200 machine. Data in figure graphs are representative of at least three independent experiments.

Digitonin permeabilization. Differential digitonin permeabilization of infected macrophages was performed as previously described³⁰.

Immunohistochemistry of mouse specimens. Paraffin-embedded specimens were deparaffinized in xylene, subjected to heat-mediated antigen-retrieval in 10 mM sodium citrate (pH 6.0), permeabilized in 0.2% Triton X-100 (Sigma), treated with mouse on mouse blocking reagent (Vector Laboratories) and blocked in 5% donkey sera. Parkin was detected using a mouse monoclonal anti-parkin antibody (Cell Signal #4211) (1:50) and an HRP-conjugated donkey anti-rabbit secondary (1:250, Jackson Immunochemicals), amplified with AB reagent (Vectastain) and detected using DAB reagent (Thermo Scientific). Images were acquired using a Zeiss Axioplan 2 microscope.

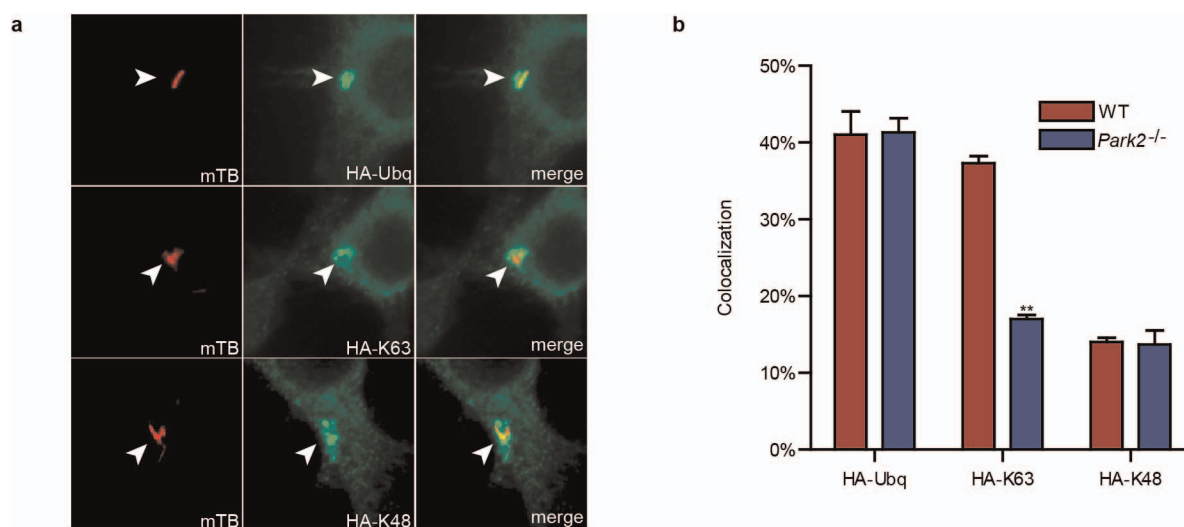
Immunohistochemistry and immunofluorescence of human specimens. Human lung biopsy specimens were obtained from patients with active tuberculosis. Paraffin-embedded specimens were deparaffinized in xylene, subjected to heat-mediated antigen-retrieval in 10 mM sodium citrate (pH 6.0), permeabilized in 0.2% Triton X-100 (Sigma) and blocked in 5% donkey sera. For immunohistochemistry, parkin was detected using a rabbit polyclonal anti-parkin antibody (Abcam) (1:25) and an HRP-conjugated donkey anti-rabbit secondary (1:250, Jackson Immunochemicals), amplified with AB reagent (Vectastain) and detected using DAB reagent (Thermo Scientific). Images were acquired using a Zeiss Axioplan 2 microscope. For immunofluorescence, parkin was identified using rabbit polyclonal anti-parkin (Abcam15954) at 1:25 and an HRP conjugated donkey anti-rabbit secondary (1:100, Jackson Immunochemicals) followed by amplification with tyramide (1:50, Perkin Elmer). TB was identified using guinea pig anti-TB (1:25, BEI Resources, NR-13818) and an Alexa488 conjugated donkey anti-guinea-pig secondary (1:100, Jackson Immunochemicals). Images were acquired using a Leica TCS SP5 confocal microscope. All human tissue specimens were obtained with consent. This study of human tissue specimens has been exempted under 45 CFR 46.101(b) and was approved by the Institutional Review Board at the University of Texas Southwestern Medical Center.

36. Goldberg, M. S. *et al.* Parkin-deficient mice exhibit nigrostriatal deficits but not loss of dopaminergic neurons. *J. Biol. Chem.* **278**, 43628–43635 (2003).
37. Ohl, Y. M. *et al.* Mycobacterium tuberculosis MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. *Cell Host Microbe* **7**, 210–220 (2010).
38. Thibault, S. T. *et al.* A complementary transposon tool kit for *Drosophila melanogaster* using *P* and piggyBac. *Nature Genet.* **36**, 283–287 (2004).
39. Ayres, J. S. & Schneider, D. S. A signaling protease required for melanization in *Drosophila* affects resistance and tolerance of infections. *PLoS Biol.* **6**, e305 (2008).



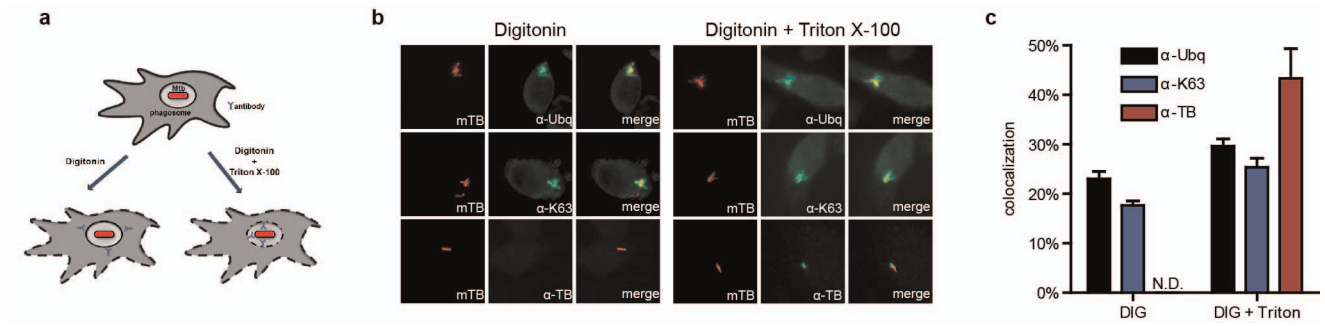
Extended Data Figure 1 | Quantification of parkin co-localization and effect of LRSAM1 knockdown in BMDMs. **a**, Quantification of parkin-positive *M. tuberculosis* in BMDMs from wild-type and *Park2*^{-/-} mice, from Fig. 1a. **b**, BMDMs from LC3-GFP transgenic mice were transduced with lentivirus expressing either a scrambled shRNA control (Ctrl) or shRNAs targeting either *LRSAM1* or parkin. Lentiviral transduced cells were then

infected with mCherry-expressing *M. tuberculosis* and the co-localization of GFP-LC3 and ubiquitin was quantified by immunofluorescence. **P* < 0.014, ***P* < 0.008 by Student's *t*-test **c**, Quantitative PCR with reverse transcription (RT-qPCR) expression of *LRSAM1* and parkin transcripts in lentiviral transduced cells from **a**. Data shown are expressed relative to actin expression. **P* < 0.033, ***P* < 0.0035 by Student's *t*-test.



Extended Data Figure 2 | Co-localization of HA-ubiquitin species during *M. tuberculosis* infection. **a**, Wild-type BMDMs were transduced with lentivirus expressing HA-tagged constructs of wild-type ubiquitin (WT), ubiquitin with all lysine residues mutated to arginine except for lysine 63 (K63), or ubiquitin with all lysine residues mutated to arginine except for lysine 48

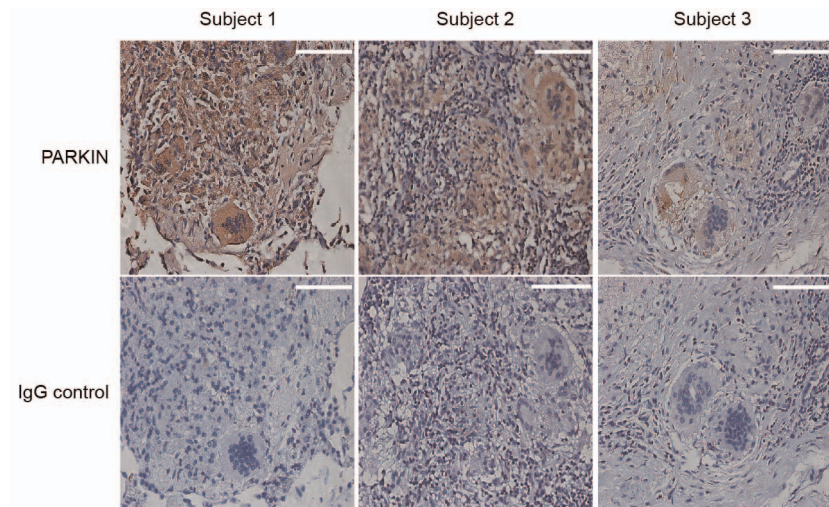
(K48). Transduced cells were then infected with mCherry-expressing *M. tuberculosis* and immunostained using anti-HA antibodies 4 h post-infection. **b**, Quantification of HA-ubiquitin co-localization with *M. tuberculosis* from **a**. ** $P < 0.001$ by Student's *t*-test.



Extended Data Figure 3 | Digitonin permeabilization of BMDMs.

a, Cartoon model showing digitonin differential permeabilization of macrophages and antibody accessibility to phagosomes. **b**, Microscopy images of wild-type BMDMs were infected with mCherry-expressing *M. tuberculosis*.

Cells were immunostained by digitonin permeabilization alone or digitonin permeabilization with Triton X-100 treatment. **c**, Quantification of ubiquitin co-localization with *M. tuberculosis* from **b**. N.D., not determined.



Extended Data Figure 4 | Immunohistochemistry analysis of parkin within human patients with active tuberculosis. Lung biopsy samples were obtained from three different human patients with active tuberculosis.

Immunohistochemistry was performed on specimens using either anti-parkin, anti-*M. tuberculosis* or an IgG control antibody. Positive cells were visualized by DAB staining. Scale bar, 100 μ m.

Swings between rotation and accretion power in a binary millisecond pulsar

A. Papitto¹, C. Ferrigno², E. Bozzo², N. Rea¹, L. Pavan², L. Burderi³, M. Burgay⁴, S. Campana⁵, T. Di Salvo⁶, M. Falanga⁷, M. D. Filipović⁸, P. C. C. Freire⁹, J. W. T. Hessels^{10,11}, A. Possenti⁴, S. M. Ransom¹², A. Riggio³, P. Romano¹³, J. M. Sarkissian¹⁴, I. H. Stairs¹⁵, L. Stella¹⁶, D. F. Torres^{1,17}, M. H. Wieringa¹⁸ & G. F. Wong^{8,14}

It is thought that neutron stars in low-mass binary systems can accrete matter and angular momentum from the companion star and be spun-up to millisecond rotational periods^{1–3}. During the accretion stage, the system is called a low-mass X-ray binary, and bright X-ray emission is observed. When the rate of mass transfer decreases in the later evolutionary stages, these binaries host a radio millisecond pulsar^{4,5} whose emission is powered by the neutron star's rotating magnetic field⁶. This evolutionary model is supported by the detection of millisecond X-ray pulsations from several accreting neutron stars^{7,8} and also by the evidence for a past accretion disc in a rotation-powered millisecond pulsar⁹. It has been proposed that a rotation-powered pulsar may temporarily switch on during periods of low mass inflow¹³ in some such systems. Only indirect evidence for this transition has hitherto been observed^{14–18}. Here we report observations of accretion-powered, millisecond X-ray pulsations from a neutron star previously seen as a rotation-powered radio pulsar. Within a few days after a month-long X-ray outburst, radio pulses were again detected. This not only shows the evolutionary link between accretion and rotation-powered millisecond pulsars, but also that some systems can swing between the two states on very short timescales.

The X-ray transient IGR J18245–2452 was first detected by INTEGRAL on 28 March 2013 and is located in the globular cluster M28 (see Supplementary Information). The X-ray luminosity of 3.5×10^{36} erg s^{−1} (0.5–10 keV), and the detection by the X-ray Telescope (XRT) on board Swift of a burst originated by a thermonuclear explosion at the surface of the compact object¹⁹, firmly classified this source as an accreting neutron star with a low-mass companion. An observation performed by XMM-Newton on 4 April 2013 revealed a coherent modulation of its X-ray emission at a period of 3.93185 ms (Figs 1 and 2). We observed delays in the pulse arrival times produced by the orbit of the neutron star around a companion star of mass $>0.17M_{\odot}$, with an orbital period of 11.0 h (see Fig. 2). The spin and orbital parameters of the source were further improved by making use of a second XMM-Newton observation, as well as two observations performed by Swift XRT (see Table 1).

Cross-referencing with the known rotation-powered radio pulsars in M28, we found that pulsar PSR J1824–2452I has ephemerides^{20,21} identical to those of the INTEGRAL X-ray source IGR J18245–2452 (see Table 1). However, the X-ray pulsations we have observed from IGR J18245–2452 are not powered by the rotation of the magnetic field, unlike the radio emission of PSR J1824–2452I. The pulse amplitude

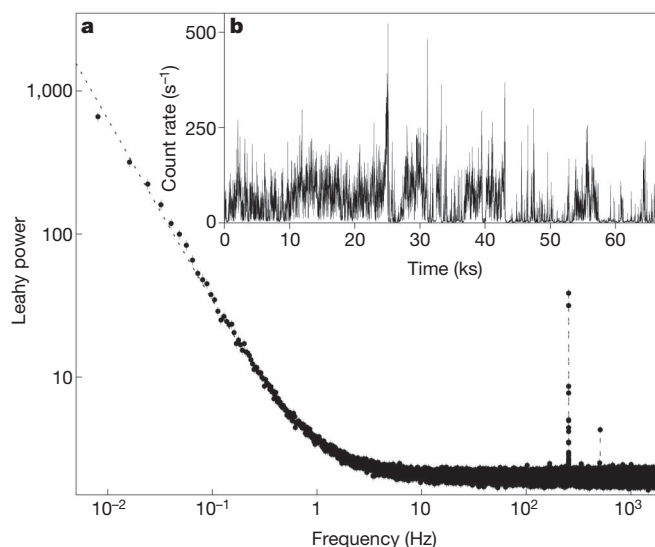


Figure 1 | Variability of the X-ray emission of IGR J18245–2452. **a**, Fourier power spectral density of the 0.5–10-keV X-ray photons observed by the EPIC pn camera on board XMM-Newton, during an observation starting on 13 April 2013, for an exposure of 67.2 ks (observation ID 0701981501). The power spectrum was obtained by sampling the light curve with a time binning of 0.236 ms, and averaging intervals 128 s in length. The times of arrival of photons were converted to the barycentre of the Solar System and to the line of nodes of the binary system hosting IGR J18245–2452, by using the parameters listed in Table 1. The peaks at 254.3 and 508.6 Hz represent the first and second harmonics of the coherent modulation of the X-ray emission of IGR J18245–2452. Considering photons observed during a 2-ks interval, not corrected for the pulsar orbital motion, the signal at the spin period of the neutron star is detected at a significance $\geq 80\sigma$. The dashed solid line is the sum of a power-law noise function, $P(f) \propto f^{-\gamma}$, with $\gamma = 1.291(4)$, and of a white noise spectrum with an average value of $1.9900(2)$ Hz^{−1}. Even considering the whole length of the time series, no break in the power-law noise could be detected at low frequencies. **b**, Light curve in the 0.5–10 keV energy band of the same observation, with a bin time of 5 s. The possibility of contamination by soft proton flares was ruled out by extracting a light curve from a background region observed by EPIC-MOS cameras far from the source. Similar properties of variability to those shown here were observed during an XMM-Newton observation starting on 3 April 2013, for an exposure of 26.7 ks (observation ID 0701981401). Error bars show $\pm 1\sigma$.

¹Institute of Space Sciences (ICE; IEEC-CSIC), Campus UAB, Faculty of Science, Torre C5, Parell, 2a Planta, E-08193 Barcelona, Spain. ²ISDC, Department of Astronomy, Université de Genève, 16 chemin d'Écogia, 16, CH-1290 Versoix, Switzerland. ³Dipartimento di Fisica, Università di Cagliari, SP Monserrato-Sestu, Km 0.7, I-09042 Monserrato, Italy. ⁴INAF – Osservatorio Astronomico di Cagliari, loc. Poggio dei Pini, strada 54, I-09012 Capoterra, Italy. ⁵INAF – Osservatorio Astronomico di Brera, Via Biancamano 46, I-23807 Merate, Lecco, Italy. ⁶Dipartimento di Fisica e Chimica, Università di Palermo, via Archirafi 36, I-90123 Palermo, Italy. ⁷International Space Science Institute, Hallerstrasse 6, CH-3012 Bern, Switzerland. ⁸University of Western Sydney, Locked Bag 1797, Penrith South DC, NSW 1797, Australia. ⁹Max-Planck-Institut für Radioastronomie, auf dem Hügel 69, 53121 Bonn, Germany. ¹⁰ASTRON, the Netherlands Institute for Radio Astronomy, Postbus 2, 7990 AA Dwingeloo, The Netherlands. ¹¹Astronomical Institute “Anton Pannekoek”, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands. ¹²National Radio Astronomy Observatory (NRAO), 520 Edgemont Road, Charlottesville, Virginia 22901, USA. ¹³INAF – Istituto di Astrofisica Spaziale e Fisica Cosmica, Via U. La Malfa 153, I-90146 Palermo, Italy. ¹⁴CSIRO Astronomy and Space Science, PO Box 76, Epping, NSW 1710, Australia. ¹⁵Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, British Columbia V6T 1Z1, Canada. ¹⁶INAF – Osservatorio Astronomico di Roma, Via di Frascati 33, I-00040 Monte Porzio Catone, Roma, Italy. ¹⁷Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain. ¹⁸CSIRO Astronomy and Space Science, Locked Bag 194, Narrabri, NSW 2390, Australia.

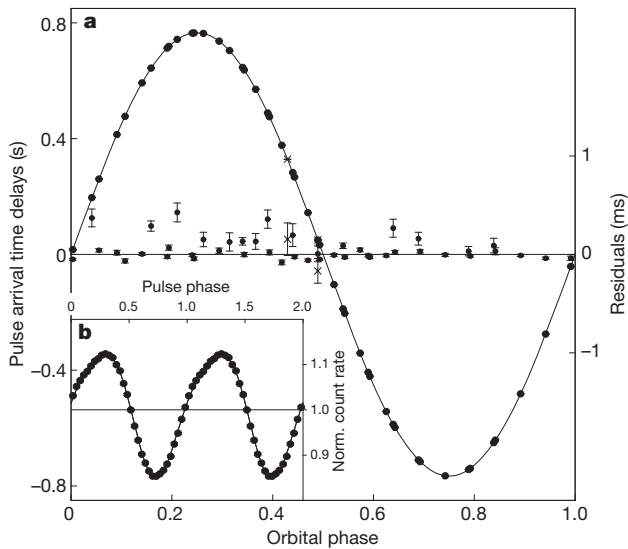


Figure 2 | Spin and orbit of IGR J18245–2452. **a**, Delays in pulse arrival time caused by the orbital motion of the neutron star (left axis) as measured by XMM-Newton during observations starting on 3 and 13 April 2013 (dots; observation IDs 0701981401 and 0701981501, with exposures of 26.7 and 67.2 ks, respectively), and by Swift during observations starting on 30 March and 7 April 2013 (crosses; observation IDs 00552369000 and 00032785005, with exposures of 0.6 and 1.6 ks, respectively). Residuals with respect to the best-fit timing solution (solid line) are also shown (right axis). Pulse profiles observed in intervals 2 ks long were modelled using $n = 12$ phase bins. The significance of each detection was assessed from the probability that the variance of each folded pulse profile was compatible with counting noise, assuming that in the absence of any signal the latter was distributed as a χ^2 variable with $(n - 1)$ degrees of freedom²⁹. Only detections with a significance larger than 3σ were considered. Delays in pulse arrival time were determined through standard methods of least-square fitting of the pulse profiles²³, using two harmonic components and considering the values measured for the fundamental frequency component. **b**, Average pulse profile sampled in 32 phase bins, accumulated over the two XMM-Newton observations (black dots), and the best-fit decomposition with two harmonics (solid line). The amplitudes of the first and second harmonics were 13.4(1)% and 1.9(1)%, respectively. Two cycles are plotted for clarity. In both panels, plotted error bars are the standard deviation of each measure.

was observed to vary in strong correlation with the X-ray flux, implying that pulsations came from a source emitting $\sim 10^{36}$ erg s^{−1} in X-rays; this value is larger by more than two orders of magnitude than the luminosity shown by the X-ray counterparts of rotation-powered radio millisecond pulsars²², whereas it agrees nicely with the X-ray output of

accretion-powered millisecond pulsars⁷. The X-ray spectrum of IGR J18245–2452 was also typical of this class, and the broad emission line observed at an energy compatible with the Fe K α transition (6.4–6.97 keV) is most easily interpreted in terms of reflection of hard X-rays by a truncated accretion disk²³ (see Supplementary Information and Supplementary Table 1). Furthermore, pulsations were detected by Swift XRT during the decay of a thermonuclear burst, after a runaway nuclear burning of light nuclei that had accreted on the neutron star surface (see Supplementary Information). Such bursts are unambiguous indicators that mass accretion is taking place¹⁹, and the oscillations observed in some of them trace the spin period of the accreting neutron star²⁴.

We derived a precise position for IGR J18245–2452 by using a Chandra image taken on 29 April 2013, while the source was fading in X-rays. Analysis of archival Chandra observations from 2008 (see Supplementary Information and Supplementary Table 2) indicate that IGR J18245–2452 was already showing variations of its X-ray luminosity by an order of magnitude, as shown in Fig. 3, suggesting that it underwent other episodes of mass accretion in the past few years. This 2008 enhancement of the X-ray emission followed the nearest previous detection of the radio pulsar, on 13 June 2008, by less than two months, indicating a very rapid transition from rotation-powered to accretion-powered activity (see Supplementary Table 3 for a summary of past observations of the source in the X-ray and radio bands). The Chandra position of IGR J18245–2452 is compatible with a variable unpulsed radio source that we detected with the Australia Compact Telescope Array on 5 April 2013, with spectral properties typical of an accreting millisecond pulsar in outburst²⁵ (see Supplementary Information).

A combination of serendipitous and target-of-opportunity observations with the Green Bank Telescope (GBT), Parkes radio telescope and Westerbork Synthesis Radio Telescope (WSRT) partly map the reactivation of IGR J18245–2452 as the radio pulsar PSR J1824–2452I (see Supplementary Information and Supplementary Table 3). No pulsed radio emission was seen in any of the three observations in April 2013, compatible with the neutron star's being in an accretion phase and inactive as a radio pulsar. However, we caution that non-detection of radio pulsations from PSR J1824–2452I can also be due to eclipsing and that the lack of observable radio pulsations does not necessarily prove the absence of an active radio pulsar mechanism^{20,22}. Radio pulsations were detected in 5 of the 13 observations conducted with GBT, Parkes and WSRT in May 2013. These observations show that the radio pulsar mechanism was active no more than a few weeks after the peak of the X-ray outburst.

In the past decade, IGR J18245–2452 has thus shown unambiguous tracers of both rotation-powered and accretion-powered activity, providing conclusive evidence for the evolutionary link between neutron stars in low-mass X-ray binaries and millisecond radio pulsars. The

Table 1 | Spin and orbital parameters of IGR J18245–2452 and PSR J1824–2452I

Parameter	IGR J18245–2452	PSR J1824–2452I
Right ascension (J2000)	18 h 24 min 32.53(4) s	
Declination (J2000)	−24° 52′ 08.6(6)″	
Reference epoch (MJD)	56386.0	
Spin period (ms)	3.931852642(2)	3.93185(1)
Spin period derivative	$<1.3 \times 10^{-17}$	
Root mean square of pulse time delays (ms)	0.1	
Orbital period (h)	11.025781(2)	11.0258(2)
Projected semimajor axis (light-seconds)	0.76591(1)	0.7658(1)
Epoch of zero mean anomaly (modified Julian date)	56395.216893(1)	
Eccentricity	$\leq 10^{-4}$	
Pulsar mass function (M_{\odot})	$2.2831(1) \times 10^{-3}$	$2.282(1) \times 10^{-3}$
Minimum companion mass (M_{\odot})	0.174(3)	0.17(1)
Median companion mass (M_{\odot})	0.204(3)	0.20(1)

Coordinates, spin, and orbital parameters of IGR J18245–2452 = PSR J1824–2452I. Celestial coordinates of IGR J18245–2452 are derived from a Chandra X-ray observation performed using the High Resolution Camera (HRC-S) on 29 April 2013 (see Fig. 3). The spin and orbital parameters of IGR J18245–2452 were derived by modelling the delays in pulse arrival time of the fundamental frequency component, as observed in the 0.5–10-keV energy band by the EPIC pn camera on board XMM-Newton, and by the XRT on board Swift (see Fig. 2 and Supplementary Information for details). The solution covers the interval between 30 March and 13 April 2013. The peak-to-peak amplitude of the fundamental varied in correlation with the observed count rate (Spearman's rank correlation coefficient $\rho = 0.79$ for 45 points, which has a probability of less than 10^{-10} if the variables are uncorrelated), with a maximum of 18%. When detected, the second harmonic has an amplitude between 2% and 3%. The minimum and median masses of the companion star were evaluated for a $1.35M_{\odot}$ mass of the neutron star, and for an inclination of the system of 90° and 60°, respectively. The spin and orbital parameters of PSR J1824–2452I were taken from ref. 20 and the Australia Telescope National Facility Pulsar Catalogue²¹, considering errors on the last significant digit there quoted. The numbers in parentheses represent the uncertainties on the respective parameter evaluated at a 1σ confidence level. Upper limits are quoted at a 3σ confidence level.

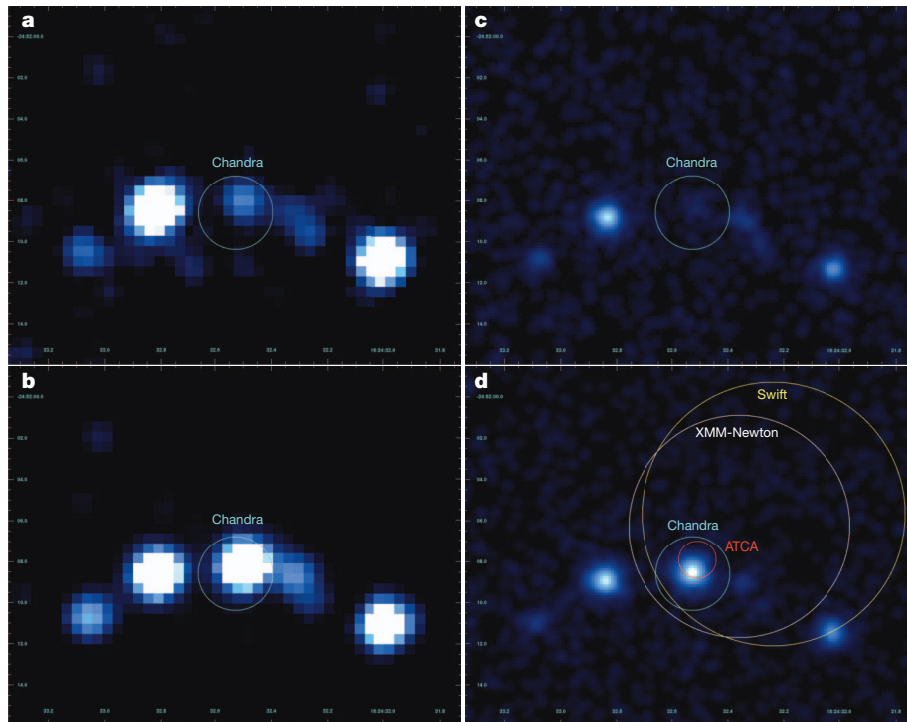


Figure 3 | Long-term X-ray variability of IGR J18245–2452. **a, b**, Chandra/ACIS-S images of the core of M28 taken during 4 August 2002 and 7 August 2008, respectively. **c, d**, Chandra/HRC-S images of the same field taken during 27 May 2006 and 29 April 2013, respectively. Images in **a** and **c** show the source in X-ray quiescence, emitting a luminosity of $\sim 1.5 \times 10^{32} \text{ erg s}^{-1}$, whereas during the observations shown in **b** and **d** the X-ray luminosity was of $\sim 2 \times 10^{33} \text{ erg s}^{-1}$ (see Supplementary Information for details). The luminosity emitted during the 2013 observation (**d**) was three orders of magnitude lower than that observed by Swift ($3.5 \times 10^{36} \text{ erg s}^{-1}$ on 30 March 2013) and XMM-Newton ($1.1 \times 10^{36} \text{ erg s}^{-1}$ on 3 April 2013) at the onset of the X-ray outburst, compatible with the source being close to the end of the accretion episode. A distance³⁰ of 5.5 kpc was considered to derive these estimates.

source swung between rotation-powered and accretion-powered states on timescales of a few days to a few months; this establishes the existence of an evolutionary phase during which a source can alternate between these two states over a timescale much shorter than the billion-year-long evolution of these binary systems, as they are spun-up by mass accretion to millisecond spin periods²⁶. It is probable that a rotation-powered pulsar switches on also during the X-ray quiescent states of other accreting millisecond pulsars^{14–18}, even if radio pulsations have not yet been detected²⁷.

The short timescales observed for the transitions between accretion-powered and rotation-powered states of IGR J18245–2452 are comparable with those typical of X-ray luminosity variations. Like other X-ray transients, IGR J18245–2452 is X-ray bright ($L_X \approx 10^{36} \text{ erg s}^{-1}$) only during a few month-long periods called ‘outbursts’; outside these episodes it spends years in an X-ray quiescent state ($L_X \lesssim 10^{32} \text{ erg s}^{-1}$). These variations are caused by swings of the mass inflow rate onto the neutron star¹³, and our findings strongly suggest that this quantity mainly regulates the transitions between accretion-powered and rotation-powered activity, which is compatible with earlier suggestions^{5,10–12}. The X-ray luminosity of IGR J18245–2452 during quiescence ($L_X \approx 10^{32} \text{ erg s}^{-1}$) implies that the rate of mass accretion was not larger than $\dot{M} \lesssim 10^{-14} M_\odot \text{ yr}^{-1}$ during such a state. The presence of millisecond radio pulsations indicates that the pulsar magnetosphere kept the plasma beyond the light cylinder radius (located at a distance of $\sim 200 \text{ km}$), despite the pressure exerted by the mass inflowing from the companion star. A pulsar magnetic field of the order of 10^8 – 10^9 G is able to satisfy this condition and to explain the quiescent X-ray luminosity in terms of the pulsar rotational power (for a typical conversion efficiency of about 1%). The irregular disappearance of the radio pulses of PSR J1824–2452I during

During the 2013 outburst, the 0.5–10-keV spectrum of IGR J18245–2452 was dominated by a power law with index ~ 1.4 , interpreted as Comptonization in an optically thin medium, of seed photons with a temperature of $\sim 0.3 \text{ keV}$. XMM-Newton observations also detected a thermal component, modelled as an accretion disc truncated at an apparent projected inner radius of $\sim 50 \text{ km}$, and a broad line, modelled with a Gaussian centred at $6.74 \pm 0.11 \text{ keV}$ and of width $1.1 \pm 0.2 \text{ keV}$, compatible with Fe K α transition (see Supplementary Information for details). The plotted error circles represent the 3σ confidence level position of IGR J18245–2452, derived by Chandra (29 April 2013), XMM-Newton EPIC-MOS (3 and 13 April 2013), Swift XRT (30 March 2013) and ATCA (5 April 2013), plotted as cyan, white, yellow and red circles, respectively.

the rotation-powered stage suggests that, during that phase, most of the matter that the companion transfers towards the neutron star is ejected by the pressure of the pulsar wind^{5,28}. A slight increase in the mass transfer rate may subsequently push the magnetosphere back inside the light cylinder¹². After a disk had sufficient time to build up, an X-ray outburst is expected to take place, as in the case of IGR J18245–2452 during the observations reported here. As the mass accretion rate decreases during the decay of the X-ray outburst, the pressure of the magnetosphere is able to at least partly sweep away the residual matter from the surroundings of the neutron star, and a rotation-powered pulsed radio emission can reactivate. Our observations prove that such transitions can take place in both directions, on a timescale shorter than expected, perhaps only a few days.

The discovery of IGR J18245–2452, swinging between rotation-powered and accretion-powered emission, represents the most stringent probe of the recycling model^{1–3}, and the existence of an unstable intermediate phase in the evolution of low-mass X-ray binaries, offering an unprecedented opportunity to study in detail the transitions between these two states.

Received 1 May; accepted 12 July 2013.

1. Bisnovatyi-Kogan, G. S. & Komberg, B. V. Pulsars and close binary systems. *Sov. Astron.* **18**, 217–221 (1974).
2. Alpar, M. A., Cheng, A. F., Ruderman, M. A. & Shaham, J. A new class of radio pulsars. *Nature* **300**, 728–730 (1982).
3. Radhakrishnan, V. & Srinivasan, G. On the origin of the recently discovered ultra-rapid pulsar. *Curr. Sci. India* **51**, 1096–1099 (1982).
4. Backer, D. C., Kulkarni, S. R., Heiles, C., Davis, M. M. & Goss, W. M. A millisecond pulsar. *Nature* **300**, 615–618 (1982).

5. Ruderman, M., Shaham, J. & Tavani, M. Accretion turnoff and rapid evaporation of very light secondaries in low-mass X-ray binaries. *Astrophys. J.* **336**, 507–518 (1989).
6. Pacini, F. Energy emission from a neutron star. *Nature* **216**, 567–568 (1967).
7. Wijnands, R. & van der Klis, M. A millisecond pulsar in an X-ray binary system. *Nature* **394**, 344–346 (1998).
8. Chakrabarty, D. & Morgan, E. H. The two-hour orbit of a binary millisecond X-ray pulsar. *Nature* **394**, 346–348 (1998).
9. Archibald, A. M. *et al.* A radio pulsar/X-ray binary link. *Science* **324**, 1411–1414 (2009).
10. Stella, L., Campana, S., Colpi, M., Mereghetti, S. & Tavani, M. Do quiescent soft X-ray transients contain millisecond radio pulsars? *Astrophys. J.* **423**, L47–L50 (1994).
11. Campana, S., Colpi, M., Mereghetti, S., Stella, L. & Tavani, M. The neutron stars of soft X-ray transients. *Astron. Astrophys. Rev.* **8**, 279–316 (1998).
12. Burderi, L. *et al.* Where may ultrafast rotating neutron stars be hidden? *Astrophys. J.* **560**, L71–L74 (2001).
13. van Paradijs, J. On the accretion instability in soft X-ray transients. *Astrophys. J.* **464**, L139–L141 (1996).
14. Burderi, L., Di Salvo, T., D'Antona, F., Robba, N. R. & Testa, V. The optical counterpart to SAX J1808.4–3658 in quiescence: evidence of an active radio pulsar? *Astron. Astrophys.* **404**, L43–L46 (2003).
15. Hartman, J. M. *et al.* The long-term evolution of the spin, pulse shape, and orbit of the accretion-powered millisecond pulsar SAX J1808.4–3658. *Astrophys. J.* **675**, 1468–1486 (2008).
16. Di Salvo, T., Burderi, L., Riggio, A., Papitto, A. & Menna, M. T. Orbital evolution of an accreting millisecond pulsar: witnessing the banquet of a hidden black widow? *Mon. Not. R. Astron. Soc.* **389**, 1851–1857 (2008).
17. Patruno, A. The accreting millisecond X-ray pulsar IGR J00291+5934: evidence for a long timescale spin evolution. *Astrophys. J.* **722**, 909–918 (2010).
18. Papitto, A. *et al.* Spin down during quiescence of the fastest known accretion-powered pulsar. *Astron. Astrophys.* **528**, A55 (2011).
19. Lewin, W. H. G., van Paradijs, J. & Taam, R. E. X-Ray bursts. *Space Sci. Rev.* **62**, 223–389 (1993).
20. Bégin, S. *A Search for Fast Pulsars in Globular Clusters*. MSc thesis, Faculty of Graduate Studies (Physics), Univ. British Columbia (2006).
21. Manchester, R. N., Hobbs, G. B., Teoh, A. & Hobbs, M. The Australia Telescope National Facility Pulsar Catalogue. *Astron. J.* **129**, 1993–2006 (2005).
22. Bogdanov, S. *et al.* Chandra X-ray observations of 12 millisecond pulsars in the globular cluster M28. *Astrophys. J.* **730**, 81 (2011).
23. Papitto, A. *et al.* XMM-Newton detects a relativistically broadened iron line in the spectrum of the ms X-ray pulsar SAX J1808.4–3658. *Astron. Astrophys.* **493**, L39–L43 (2009).
24. Chakrabarty, D. *et al.* Nuclear-powered millisecond pulsars and the maximum spin frequency of neutron stars. *Nature* **424**, 42–44 (2003).
25. Gaensler, B. M., Stappers, B. W. & Getts, T. J. Transient radio emission from SAX J1808.4–3658. *Astrophys. J.* **522**, L117–L119 (1999).
26. Bhattacharya, D. & van den Heuvel, E. P. J. Formation and evolution of binary and millisecond radio pulsars. *Phys. Rep.* **203**, 1–124 (1991).
27. Burgay, M. *et al.* A search for pulsars in quiescent soft X-ray transients. I. *Astrophys. J.* **589**, 902–910 (2003).
28. Fruchter, A. S., Stinebring, D. R. & Taylor, J. H. A millisecond pulsar in an eclipsing binary. *Nature* **333**, 237–239 (1988).
29. Leahy, D. A., Elsner, R. F. & Weisskopf, M. C. On searches for periodic pulsed emission—the Rayleigh test compared to epoch folding. *Astrophys. J.* **272**, 256–258 (1983).
30. Harris, W. E. A catalog of parameters for globular clusters in the Milky Way. *Astron. J.* **112**, 1487–1488 (1996).

Supplementary Information is available in the online version of the paper.

Acknowledgements This letter is based on ToO observations made with XMM-Newton, Chandra, INTEGRAL, Swift, ATCA, WSRT, GBT and PKS. We thank the respective directors and operation teams for their support. Work was done in the framework of grants AYA2012-39303, SGR2009-811 and iLINK2011-0303, and with the support of CEA/Irfu, IN2P3/CNRS and CNES (France), INAF (Italy), NWO (The Netherlands) and NSERC (Canada). A.Pa. is supported by a Juan de la Cierva Research Fellowship. A.R. acknowledges Sardinia Regional Government for financial support (P.O.R. Sardegna ESF 2007-13). D.F.T. was additionally supported by a Friedrich Wilhelm Bessel Award of the Alexander von Humboldt Foundation. L.P. thanks the Société Académique de Genève and the Swiss Society for Astrophysics and Astronomy. We acknowledge the use of data supplied by the UK Swift Science Data Centre at the University of Leicester. A.Pa. thanks S. Giannetti, D. Lai, R. V. E. Lovelace, M. M. Romanova and T. Tauris for discussions, and S. D. Wolf for operational support.

Author Contributions A.Pa., C.F. and E.B. collected and analysed XMM-Newton data. A.Pa. and C.F. detected the pulsar in XMM-Newton data and derived its orbital solution. A.Pa. discovered the equivalence of its parameters with a radio pulsar, the thermonuclear burst and the burst oscillations. N.R. analysed Chandra data, detecting the X-ray quiescent counterpart of the source and past accretion events. L.P., M.H.W., M.D.F. and G.F.W. analysed ATCA data. E.B., S.C., P.R., A.Pa. and A.R. analysed Swift data. E.B. and C.F. analysed INTEGRAL data. J.W.T.H. analysed WSRT data. M.B. and J.M.S. analysed PKS data. J.W.T.H., S.M.R., A.Po., I.H.S. and P.C.C.F. analysed GBT data. A.R. provided software tools. A.Pa., N.R. and J.W.T.H. wrote the manuscript, with significant contribution by all the authors in interpreting the results and editing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.Pa. (papitto@ice.csic.es).

Observation of dipolar spin-exchange interactions with lattice-confined polar molecules

Bo Yan^{1,2}, Steven A. Moses^{1,2}, Bryce Gadway^{1,2}, Jacob P. Covey^{1,2}, Kaden R. A. Hazzard^{1,2}, Ana Maria Rey^{1,2}, Deborah S. Jin^{1,2} & Jun Ye^{1,2}

With the production of polar molecules in the quantum regime^{1,2}, long-range dipolar interactions are expected to facilitate understanding of strongly interacting many-body quantum systems and to realize lattice spin models³ for exploring quantum magnetism. In ordinary atomic systems, where contact interactions require wavefunction overlap, effective spin interactions on a lattice can be mediated by tunnelling, through a process referred to as superexchange; however, the coupling is relatively weak and is limited to nearest-neighbour interactions^{4,5}. In contrast, dipolar interactions exist even in the absence of tunnelling and extend beyond nearest neighbours. This allows coherent spin dynamics to persist even for gases with relatively high entropy and low lattice filling. Measured effects of dipolar interactions in ultracold molecular gases have been limited to the modification of inelastic collisions and chemical reactions^{6,7}. Here we use dipolar interactions of polar molecules pinned in a three-dimensional optical lattice to realize a lattice spin model. Spin is encoded in rotational states of molecules that are prepared and probed by microwaves. Resonant exchange of rotational angular momentum between two molecules realizes a spin-exchange interaction. The dipolar interactions are apparent in the evolution of the spin coherence, which shows oscillations in addition to an overall decay of the coherence. The frequency of these oscillations, the strong dependence of the spin coherence time on the lattice filling factor and the effect of a multipulse sequence designed to reverse dynamics due to two-body exchange interactions all provide evidence of dipolar interactions. Furthermore, we demonstrate the suppression of loss in weak lattices due to a continuous quantum Zeno mechanism⁸. Measurements of these tunnelling-induced losses allow us to determine the lattice filling factor independently. Our work constitutes an initial exploration of the behaviour of many-body spin models with direct, long-range spin interactions and lays the groundwork for future studies of many-body dynamics in spin lattices.

Long-range and spatially anisotropic dipole-dipole interactions permit new approaches for the preparation and exploration of strongly correlated quantum matter that exhibits intriguing phenomena such as quantum magnetism, exotic superfluidity and topological phases^{9–13}. Ultracold gases of polar molecules provide highly controllable, long-lived and strongly interacting dipolar systems and have recently attracted intense scientific interest. Samples of fermionic ⁴⁰K⁸⁷Rb polar molecules, with an electric dipole moment of 0.57 D (ref. 1; 1 D = 3.336 × 10^{–30} C m), have been prepared near the Fermi temperature, and all degrees of freedom (electronic, vibrational, rotational, hyperfine and external motion) can be controlled at the level of single quantum states^{7,14,15}.

The surprising discovery of bimolecular chemical reactions of KRb at ultralow temperatures^{2,6,7} seemed to be a major challenge in creating novel quantum matter. However, the molecules' motion, and, consequently, their reactions, can be fully suppressed in a three-dimensional (3D) optical lattice, where relatively long lifetimes (>25 s) have been observed¹⁵. The long-range dipolar interaction can then play the dominant part in the

dynamics of the molecular internal degrees of freedom, for example by exchanging two neighbouring molecules' rotational states. With spin encoded in the rotational states of the molecule, these dipolar interactions give rise to spin-exchange interactions, analogous to those that are important in quantum magnetism and high-temperature superconductivity¹⁶. In a 3D lattice, where each molecule is surrounded by many neighbouring sites, this system represents an intriguing many-body quantum spin system in which excitations can have strong correlations even at substantially less than unit lattice filling¹⁷.

Several features distinguish the interactions in a molecular spin model from those observed in ultracold atomic systems. For the superexchange interaction of atoms in optical lattices^{4,5}, the short-range nature of the interparticle interactions necessitates a second-order perturbative process to occur in the tunnelling of atoms between lattice sites. Hence, the energy scale of the superexchange interaction decreases exponentially with lattice depth. This spin-motion coupling limits superexchange to nearest-neighbour interactions and requires extremely low temperature and entropy.

In contrast, long-range dipolar interactions decay with separation, r , as $1/r^3$, and interactions beyond nearest neighbours are significant. This long-range interaction allows exploration of coherent spin dynamics in very deep lattices where the molecules' translational motion is frozen and where the absence of tunnelling would preclude the superexchange interactions of atoms. We note that the dipolar interaction is also different from that of electrons, for which an effective spin interaction arises due to the spin-independent Coulomb interaction and the exchange symmetry of the fermionic electrons. In contrast, the dipolar interaction is a direct spin-spin interaction that does not require any wavefunction overlap. In addition to polar molecules, ultracold systems such as magnetic atoms^{10,18–20} and trapped ions^{21,22} are candidates for realizing coherent, controllable spin models with power-law interactions; however, spin-exchange interactions have yet to be created and observed in these systems. In Rydberg atoms, Förster resonances involving multiple Rydberg states have been observed, albeit with short coherence times²³.

The molecular rotational states $|N, m_N\rangle$, where N is the principal quantum number and m_N is the projection onto the quantization axis, are the focus of our current investigation of a dipolar spin system. In general, an external d.c. electric field induces a dipole moment in the laboratory frame by mixing opposite-parity rotational states. However, even in the absence of a d.c. electric field, dipolar interactions can be established using a microwave field to create a coherent superposition between two rotational states²⁴, labelled $|\uparrow\rangle$ and $|\downarrow\rangle$. In addition, a microwave field can probe the coherent spin dynamics due to dipolar interactions.

In the absence of an applied electric field, two-level polar molecules trapped in a strong 3D lattice (Fig. 1a) can be described using a spin-1/2 lattice model with the interaction Hamiltonian^{12,17,24}

$$H = \frac{J_{\perp}}{2} \sum_{i>j} V_{\text{dd}}(\mathbf{r}_i - \mathbf{r}_j) (S_i^+ S_j^- + S_i^- S_j^+) \quad (1)$$

¹JILA, National Institute of Standards and Technology and University of Colorado, Boulder, Colorado 80309, USA. ²Department of Physics, University of Colorado, Boulder, Colorado 80309, USA.

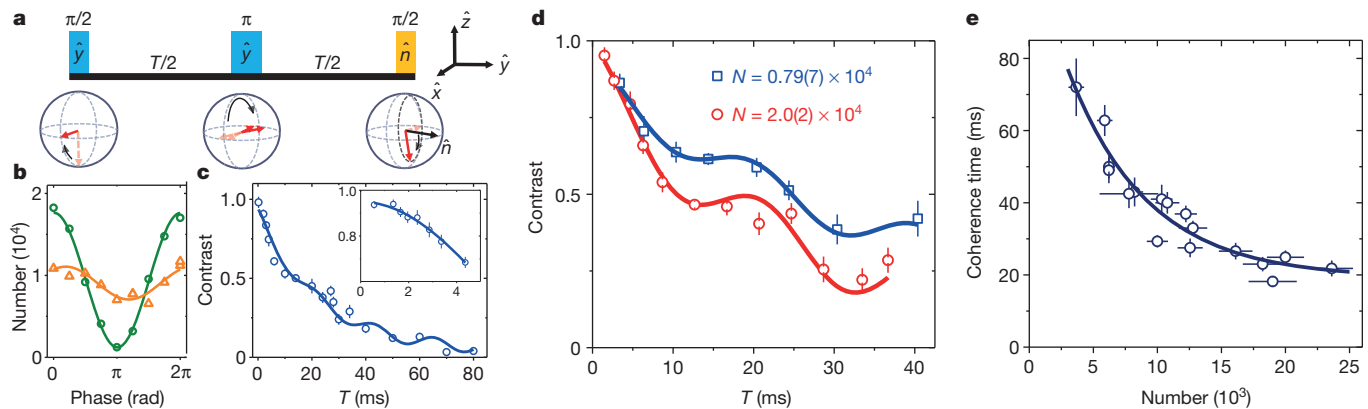


Figure 2 | Coherent spin dynamics of polar molecules. **a**, A $(\pi/2)_y$ -pulse initializes the molecules in a coherent superposition of rotational states. A spin-echo pulse sequence is used to correct for effects arising from single-particle inhomogeneities across the sample, such as residual light shifts. **b**, The phase of the final $\pi/2$ -pulse is scanned (corresponding to rotations around a variable axis, \hat{n}) to obtain a Ramsey fringe. Two fringes are shown, corresponding to the short (green circles) and intermediate timescales (orange triangles). **c**, The contrast of the Ramsey fringe is measured as a function of interrogation time.

but τ depends on the number of molecules, or filling fraction. This is a signature of many-body interactions, and the observation agrees with theoretical calculations using a cluster expansion (Methods).

We observe oscillation frequencies in the range 48 ± 2 Hz for molecule numbers in the lattice that vary by threefold. The fact that this frequency is consistent with the largest nearest-neighbour interaction energy, $|J_{\perp}/2\hbar| = 52$ Hz, supports the conclusion that the contrast oscillations come from nearest-neighbour dipole–dipole interactions. Because this frequency is determined by the lattice geometry and the dipole matrix element, it does not depend on the lattice filling factor. We also confirm that the oscillation frequency does not depend on the lattice depth from $20E_r$ to $50E_r$ (Methods). For the coherence time, we observe a strong dependence on the filling factor (Fig. 2e). Density dependence is a classic signature of interaction effects, and we conclude that the coherence time in the deep lattice is limited by dipole–dipole interactions. For higher filling factors, the increasing probability that molecules have multiple neighbours means that more spin-exchange frequencies will contribute to the signal, which leads to faster dephasing.

Multipulse sequences, as well as single spin-echo pulses, are examples of dynamical decoupling, which is widely used in NMR²⁷ and

quantum information processing²⁸ to remove dephasing and extend coherence times. Although a spin-echo pulse cannot mitigate the contrast decay that arises from dipole–dipole interactions, a multipulse sequence can. In particular, the pulse sequence²⁷ shown in Fig. 3a is designed to remove dephasing due to two-particle dipolar interactions (Methods). Analogous to how a spin-echo pulse works, this pulse sequence swaps the eigenstates of the dipolar interaction Hamiltonian (equation (1)) for two isolated particles to allow for subsequent rephasing.

Figure 3b summarizes the decay of the Ramsey contrast for three different pulse sequences. With a simple two-pulse Ramsey sequence (with no spin-echo pulse), the coherence time of the system is very short, with the fringe contrast decaying within 1 ms (Fig. 3b, triangles). With the addition of a single spin-echo pulse, the single-particle dephasing time can be extended to ~ 80 ms (measured for our lowest molecular density). However, this coherence time is reduced drastically with increasing molecule number in the lattice, and we observe oscillations in the contrast signal (Fig. 3b, circles). When we apply the multipulse sequence, the oscillations in the contrast are suppressed and the data fit well a simple exponential decay with a coherence time slightly longer than that of the spin-echo case (Fig. 3b, squares). The

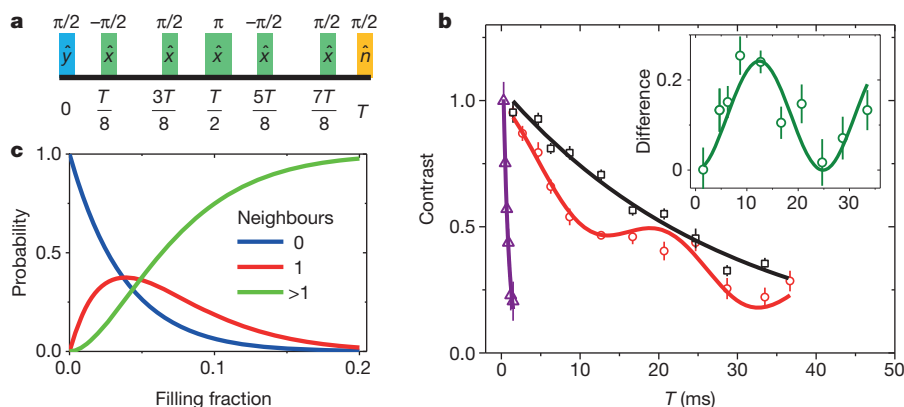


Figure 3 | Multipulse sequence and decoupling of pairwise dipolar interactions. **a**, The multipulse sequence is designed to suppress both single-particle dephasing and the effect of pairwise dipole–dipole interactions. **b**, The contrast decay is displayed as a function of time for three different pulse sequences. Without a spin-echo pulse, single-particle inhomogeneities result in a Ramsey coherence time of ~ 1 ms (triangles). The spin-echo pulse effectively removes the single-particle dephasing, such that spin-exchange interactions

play the dominant role in the contrast decay (circles). The multipulse sequence suppresses the contrast oscillations and slightly improves the coherence time (squares). Inset, the difference in contrast between the multipulse sequence and the spin-echo case shows oscillations. Error bars, 1 s.d. **c**, The probability of a particular molecule having zero, one or more than one neighbours (within the cube shown in Fig. 1c) is plotted as a function of a uniform lattice filling factor.

differences in the measured contrast oscillations and decay for the usual spin-echo and multipulse sequence highlight the spin-exchange dynamics driven by pairwise dipolar interactions (Fig. 3b, inset).

To understand the dynamics of this spin system, a key ingredient is the filling fraction of molecules in the 3D lattice, because the Ramsey contrast decay depends sensitively on the molecular density (Fig. 2e). Figure 3c shows the probabilities of a particular molecule having zero, one or more than one neighbours. The probability of having two or more neighbours is non-negligible even for relatively low fillings. The contrast oscillation is dominated by contributions from pairs of molecules, whereas interactions of multiple molecules contribute mainly to the contrast decay.

To provide an independent determination of the filling fraction, we have measured tunnelling-induced loss at reduced lattice depths. Molecules are initially prepared in $|\downarrow\rangle$ in a $40E_r$ lattice. For our fermionic molecules, the chemical reaction rate is much larger between molecules in distinguishable internal states². Moreover, Pauli blocking strongly suppresses the tunnelling of molecules in the same spin state into the same lattice site. Therefore, we create a 50:50 incoherent spin mixture of $|\downarrow\rangle$ and $|\uparrow\rangle$ by applying a $\pi/2$ -pulse and waiting 50 ms. We then quickly (within 1 ms) lower the lattice depth along only a single direction (\hat{Y} , as shown in Fig. 4a) to allow tunnelling and loss due to on-site chemical reactions^{2,6,7}. We then measure the remaining number of molecules in the $|\downarrow\rangle$ state as a function of the holding time. Figure 4b shows example loss curves for two different lattice depths along \hat{Y} .

In our system, the on-site loss rate, Γ_0 , is proportional to the chemical reaction rate between the $|0,0\rangle$ and $|1,-1\rangle$ molecules⁷:

$$\Gamma_0 = \beta \int |W(X,Y,Z)|^4 dX dY dZ \quad (2)$$

Here $\beta = 9.0(4) \times 10^{-10} \text{ cm}^3 \text{ s}^{-1}$ (parenthetical error, s.d.) is the two-body loss coefficient (Methods) and $W(X,Y,Z)$ is the ground-band Wannier function. We can modify Γ_0 by changing the lattice depth;

however, for our measurements, the system always remains in the strongly interacting regime, in which $\Gamma_0 \gg J_t/\hbar$, where J_t is the tunnelling amplitude. This is the regime of the continuous quantum Zeno effect^{8,29}, where dissipation in the form of measurement or loss can lead to suppression of coherent processes such as tunnelling. Thus, increasing Γ_0 actually decreases the effective two-body loss rate between neighbouring molecules, which is given by

$$\Gamma_{\text{eff}} = \frac{2(J_t/\hbar)^2}{\Gamma_0} \quad (3)$$

The number of $|\downarrow\rangle$ molecules, $N_{\downarrow}(t)$, can then be described with a two-body loss equation

$$\frac{dN_{\downarrow}(t)}{dt} = -\frac{\kappa}{N_{\downarrow,0}} N_{\downarrow}(t)^2 \quad (4)$$

where $N_{\downarrow,0}$ is the initial number of $|\downarrow\rangle$ molecules and the loss rate coefficient is given by³⁰ $\kappa = 4q\Gamma_{\text{eff}}g_{\downarrow\uparrow}^{(2)}n_{\downarrow,0}$. Here $2n_{\downarrow,0} = n_0$ is the initial filling fraction in the lattice, $q = 2$ is the number of nearest-neighbour sites in our one-dimensional tunnelling geometry and $g_{\downarrow\uparrow}^{(2)}$ is the correlation function of different spin states for nearest-neighbouring sites i and j : $g_{\downarrow\uparrow}^{(2)} = \langle \hat{n}_i \hat{n}_j - 4\vec{S}_i \cdot \vec{S}_j \rangle / \langle \hat{n}_i \rangle^2$, with \hat{n}_i the number operator at site i and \vec{S}_i the spin-1/2 vector operator. In our case, we assume that the molecules are initially randomly distributed in the $|\downarrow\rangle$ and $|\uparrow\rangle$ states, so that $g_{\downarrow\uparrow}^{(2)} = 1$. Because the redistribution of molecules due to losses and tunnelling can modify $g_{\downarrow\uparrow}^{(2)}$, we fit the data to the solution of equation (4) for short times, where the number has changed by less than 50%.

We verify the scaling of the continuous quantum Zeno effect by measuring the dependence of the loss rate, κ , on Γ_0 and J_t . To study the dependence on Γ_0 , we set the lattice depth along \hat{Y} to be $5.4(4)E_r$, which fixes J_t , and then increase the lattice depths along the \hat{X} and \hat{Z}

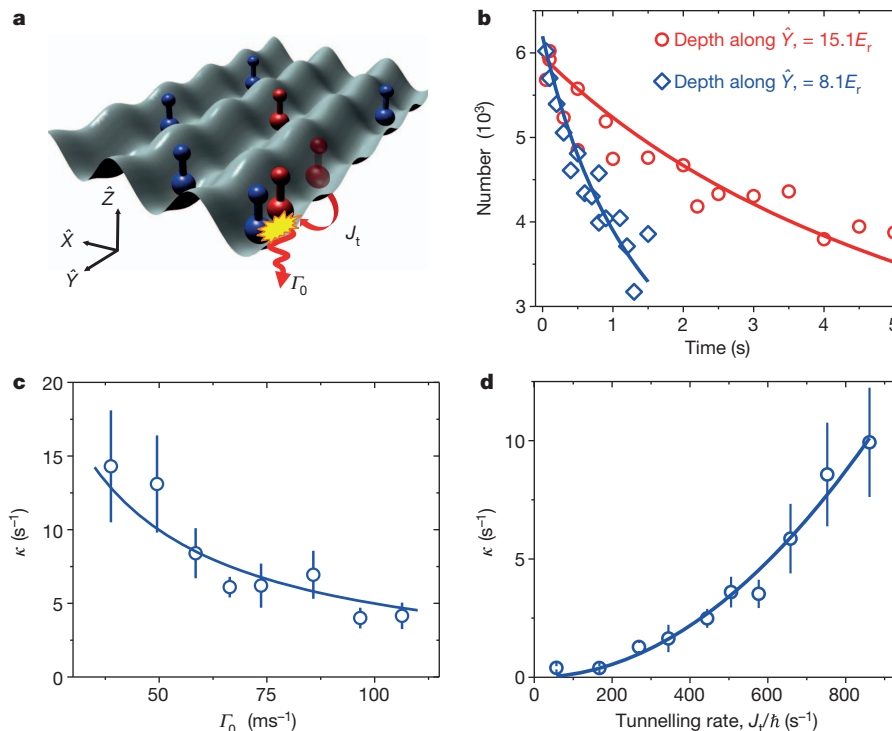


Figure 4 | Quantum Zeno effect for polar molecules in a 3D lattice. **a**, The lattice depths along \hat{X} and \hat{Z} are kept at $40E_r$, whereas the lattice depth along \hat{Y} is reduced to allow tunnelling along the \hat{Y} direction at a rate J_t/\hbar . Once two molecules in different spin states tunnel to the same site, they are lost owing to chemical reactions occurring at rate Γ_0 . **b**, Number of $|\downarrow\rangle$ molecules versus time

is shown for lattice depths along \hat{Y} of $8.1E_r$ and $15.1E_r$. **c**, The number loss rate, κ , has a $1/\Gamma_0$ dependence (fit shown), which is consistent with the quantum Zeno effect. **d**, The number loss rate, κ , has a J_t^2 dependence (fit shown), as predicted from the quantum Zeno effect. Error bars, 1 s.d.

directions. This compresses the wave function $W(X, Y, Z)$ at each lattice site, and thus increases Γ_0 . As expected for the quantum Zeno regime, the measured κ decreases as Γ_0 increases, and the data are consistent with $\kappa \propto 1/\Gamma_0$ (Fig. 4c). To study the dependence on J , we vary the lattice depth along \hat{Y} , while adjusting the \hat{X} and \hat{Z} lattice depths to keep Γ_0 fixed. As shown in Fig. 4d, the measured κ exhibits a quadratic dependence on J as predicted by equation (3). For these loss rate measurements, all parameters are known except the initial filling fraction, n_0 . From measurements of the loss rate at several lattice depths, our simple model gives $n_0 = 25(5)\%$ for 2×10^4 molecules. However, a more complete theory that incorporates interaction-modified Wannier functions in the lattice will lower this value. We note that calculations of the Ramsey fringe contrast decay using a cluster expansion give an estimated n_0 of $\sim 10\%$ (Methods), which is more consistent with an estimation of the filling based on direct imaging.

Although it is desirable to increase the lattice filling to explore interesting phases such as quantum magnetism or exotic superfluidity, we have seen that the modest filling factors achieved in our experiment already enable the observation of dipolar interaction effects in a 3D lattice spin model. Furthermore, this work prepares us for the study of non-trivial dynamic processes such as many-body localization of spin excitations. Adding an external electric field would further increase the variety of spin models that can be realized with this system.

METHODS SUMMARY

We begin with $\sim 1 \times 10^5$ ^{87}Rb atoms and 2.5×10^5 ^{40}K atoms in a far off-resonance dipole trap at 1,064 nm. The trap frequencies are 25 Hz radially and 185 Hz axially for Rb, where the axial direction is along \hat{Z} . The Rb gas is a Bose–Einstein condensate with $T/T_c \approx 0.5$ and the K Fermi gas is at $T/T_F \approx 0.5$, where T_c is the transition temperature for the condensate and T_F is the Fermi temperature. We smoothly ramp on a 3D lattice over 100 ms to a final depth of $40E_r$ (16 and 7 recoil energies for Rb and K atoms, respectively). The \hat{X} and \hat{Y} lattice beams have waists of 200 μm and the \hat{Z} beam has a waist of 250 μm . The lattice depth is calibrated with parametric heating of the molecular gas²⁵ and has an estimated uncertainty of 5%. After turning on the lattice, we lower the intensity of the dipole trap to zero in 50 ms, and then ramp a magnetic field from 54.89 to 54.59 mT in 1 ms to create weakly bound KRb Feshbach molecules. We then use two-photon stimulated Raman adiabatic passage to transfer the Feshbach molecules to the rovibrational ground state. The unpaired Rb and K atoms are removed using resonant light scattering. From band-mapping measurements, we find that the fraction of molecules in higher bands is consistent with zero within our detection limit of 5%. To measure the number of ground-state molecules in the lattice, we reverse the two-photon stimulated Raman adiabatic passage to recreate Feshbach molecules, and then take an absorption image using light resonant with the K cycling transition.

Full Methods and any associated references are available in the online version of the paper.

Received 17 May; accepted 16 July 2013.

Published online 18 September 2013.

1. Ni, K.-K. *et al.* A high phase-space-density gas of polar molecules. *Science* **322**, 231–235 (2008).
2. Ospelkaus, S. *et al.* Quantum-state controlled chemical reactions of ultracold potassium-rubidium molecules. *Science* **327**, 853–857 (2010).
3. Micheli, A., Brennen, G. K. & Zoller, P. A toolbox for lattice-spin models with polar molecules. *Nature Phys.* **2**, 341–347 (2006).
4. Trotzky, S. *et al.* Time-resolved observation and control of superexchange interactions with ultracold atoms in optical lattices. *Science* **319**, 295–299 (2008).
5. Greif, D., Uehlinger, T., Jotzu, G., Tarruell, L. & Esslinger, T. Short-range quantum magnetism of ultracold fermions in an optical lattice. *Science* **340**, 1307–1310 (2013).

6. Ni, K.-K. *et al.* Dipolar collisions of polar molecules in the quantum regime. *Nature* **464**, 1324–1328 (2010).
7. de Miranda, M. H. G. *et al.* Controlling the quantum stereodynamics of ultracold bimolecular reactions. *Nature Phys.* **7**, 502–507 (2011).
8. Syassen, N. *et al.* Strong dissipation inhibits losses and induces correlations in cold molecular gases. *Science* **320**, 1329–1331 (2008).
9. Carr, L. D., DeMille, D., Krets, R. V. & Ye, J. Cold and ultracold molecules: science, technology and applications. *New J. Phys.* **11**, 055049 (2009).
10. Lahaye, T., Menotti, C., Santos, L., Lewenstein, M. & Pfau, T. The physics of dipolar bosonic quantum gases. *Rep. Prog. Phys.* **72**, 126401 (2009).
11. Potter, A. C., Berg, E., Wang, D.-W., Halperin, B. I. & Demler, E. Superfluidity and dimerization in a multilayered system of fermionic polar molecules. *Phys. Rev. Lett.* **105**, 220406 (2010).
12. Gorshkov, A. V. *et al.* Tunable superfluidity and quantum magnetism with ultracold polar molecules. *Phys. Rev. Lett.* **107**, 115301 (2011).
13. Yao, N. Y. *et al.* Topological flat bands from dipolar spin systems. *Phys. Rev. Lett.* **109**, 266804 (2012).
14. Ospelkaus, S. *et al.* Controlling the hyperfine state of rovibronic ground-state polar molecules. *Phys. Rev. Lett.* **104**, 030402 (2010).
15. Chotia, A. *et al.* Long-lived dipolar molecules and Feshbach molecules in a 3D optical lattice. *Phys. Rev. Lett.* **108**, 080405 (2012).
16. Lee, P. A., Nagaosa, N. & Wen, X.-G. Doping a Mott insulator: physics of high-temperature superconductivity. *Rev. Mod. Phys.* **78**, 17–85 (2006).
17. Hazzard, K. R. A., Manmana, S. R., Foss-Feig, M. & Rey, A. M. Far-from-equilibrium quantum magnetism with ultracold polar molecules. *Phys. Rev. Lett.* **110**, 075301 (2013).
18. Lu, M., Burdick, N. Q. & Lev, B. L. Quantum degenerate dipolar Fermi gas. *Phys. Rev. Lett.* **108**, 215301 (2012).
19. Aikawa, K. *et al.* Bose–Einstein condensation of erbium. *Phys. Rev. Lett.* **108**, 210401 (2012).
20. de Paz, A. *et al.* Resonant demagnetization of a dipolar BEC in a 3D optical lattice. *Phys. Rev. A* **87**, 051609(R) (2013).
21. Britton, J. W. *et al.* Engineered two-dimensional Ising interactions in a trapped-ion quantum simulator with hundreds of spins. *Nature* **484**, 489–492 (2012).
22. Islam, R. *et al.* Emergence and frustration of magnetism with variable-range interactions in a quantum simulator. *Science* **340**, 583–587 (2013).
23. Nipper, J. *et al.* Atomic pair-state interferometer: controlling and measuring an interaction-induced phase shift in Rydberg-atom pairs. *Phys. Rev. X* **2**, 031011 (2012).
24. Barnett, R., Petrov, D., Lukin, M. & Demler, E. Quantum magnetism with multicomponent dipolar molecules in an optical lattice. *Phys. Rev. Lett.* **96**, 190401 (2006).
25. Neyenhuis, B. *et al.* Anisotropic polarizability of ultracold polar KRb molecules. *Phys. Rev. Lett.* **109**, 230403 (2012).
26. Martin, M. J. *et al.* A quantum many-body spin system in an optical lattice clock. *Science* **341**, 632–636 (2013).
27. Waugh, J. S., Huber, L. M. & Haeberlen, U. Approach to high-resolution NMR in solids. *Phys. Rev. Lett.* **20**, 180–182 (1968).
28. Maurer, P. C. *et al.* Room-temperature quantum bit memory exceeding one second. *Science* **336**, 1283–1286 (2012).
29. Itano, W. M., Heinzen, D. J., Bollinger, J. J. & Wineland, D. J. Quantum Zeno effect. *Phys. Rev. A* **41**, 2295–2300 (1990).
30. Baur, S. K. & Mueller, E. J. Two-body recombination in a quantum-mechanical lattice gas: entropy generation and probing of short-range magnetic correlations. *Phys. Rev. A* **82**, 023626 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank B. Zhu, M. Foss-Feig, G. Quémener and M. Lukin for discussions. We acknowledge funding for this work from the NIST, NSF, AFOSR-ARO (MURI), ARO, DOE and ARO-DARPA-OLE. S.A.M. is supported by an NDSEG Graduate Fellowship. B.G. and K.R.A.H. are National Research Council postdoctoral fellows. K.R.A.H. and A.M.R. thank the KITP for hospitality.

Author Contributions The experimental work and data analysis were carried out by B.Y., S.A.M., B.G., J.P.C., D.S.J. and J.Y. Theoretical modelling and calculations were done by K.R.A.H. and A.M.R. All authors discussed the results and contributed to the preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.S.J. (jin@jila1.colorado.edu) or J.Y. (ye@jila.colorado.edu).

METHODS

Preparation of molecules in a 3D optical lattice. We begin with $\sim 1 \times 10^5$ ^{87}Rb atoms and 2.5×10^5 ^{40}K atoms in a far off-resonance dipole trap at 1,064 nm. The trapping frequencies are 25 Hz radially and 185 Hz axially for Rb, where the axial direction is along \hat{Z} . The Rb gas is a Bose–Einstein condensate with $T/T_c \approx 0.5$, and the K Fermi gas is at $T/T_F \approx 0.5$, where T_c is the transition temperature for the condensate and T_F is the Fermi temperature. We smoothly ramp on a 3D lattice over 100 ms to a final depth of $40E_r$ (16 and 7 recoil energies for Rb and K atoms, respectively). The \hat{X} and \hat{Y} lattice beams have waists of 200 μm and the \hat{Z} beam has a waist of 250 μm . We calibrate the lattice depth through parametric heating of the molecular gas²⁵, which results in an estimated uncertainty of 5%. After turning on the lattice, we lower the intensity of the dipole trap to zero in 50 ms, and then ramp a magnetic field from 54.89 to 54.59 mT in 1 ms to create weakly bound KRB Feshbach molecules. We then use two-photon stimulated Raman adiabatic passage to transfer the Feshbach molecules to the rovibrational ground state. The unpaired Rb and K atoms are removed using resonant light scattering. After molecules are created in the lattice, we can perform band-mapping measurements by turning off the lattice in 1 ms. We find that the fraction of molecules in higher bands is consistent with zero within our detection limit of 5%. To measure the number of ground-state molecules in the lattice, we reverse the two-photon stimulated Raman adiabatic passage to recreate Feshbach molecules, and then take an absorption image using light resonant with the K cycling transition.

Differential light shift in a 3D optical lattice. Molecules have complex internal structure; hence, there are a number of different approaches to finding a ‘magic’ trap that matches the polarizabilities of two different internal states. Because the polarizability of molecules is anisotropic, tuning the angle between the quantization axis and the polarization of the light field can change the polarizabilities²⁵. For a 3D lattice, there are three different polarization vectors. The lattice geometry in our experiment is shown in Fig. 1c. We choose the \hat{X} and \hat{Y} lattice beams to have their polarizations along the horizontal plane, at angles of $\pm 45^\circ$ relative to the magnetic field. The \hat{Z} lattice polarization is the same as that of the \hat{X} lattice. Following our previous work²⁵, the energy shifts for the $|1, 0\rangle$, $|1, -1\rangle$ and $|1, 1\rangle$ states are determined by finding the eigenvalues of the Hamiltonian

$$H = -\alpha(45^\circ)I_X - \alpha(-45^\circ)I_Y - \alpha(45^\circ)I_Z + \text{diag}(\epsilon_1, \epsilon_2, \epsilon_3) \quad (5)$$

where I_X , I_Y and I_Z are respectively the intensities of lattice beams along the \hat{X} , \hat{Y} and \hat{Z} directions, α is the polarizability matrix defined in ref. 25, and ϵ_1 , ϵ_2 and ϵ_3 are respectively the bare energies for $|0, 0\rangle$, $|1, -1\rangle$ and $|1, 1\rangle$.

Supplementary Fig. 1 shows the differential light shift (with respect to $|0, 0\rangle$) of $|1, 0\rangle$, $|1, -1\rangle$ and $|1, 1\rangle$ as a function of the lattice depth. The state $|1, -1\rangle$ has the smallest intensity dependence, which corresponds to minimal inhomogeneity due to spatial variations of the light shift. The inset shows an expanded plot for $|1, -1\rangle$. The red points are the experimentally measured transition frequencies for different lattice depths, which agree well with theory. When the lattice depth is $\sim 40E_r$ in each direction, the differential light shift is zero. We measure the transition frequency between $|0, 0\rangle$ and $|1, -1\rangle$ in a $40E_r$ lattice to be 2.22778338(8) GHz, which agrees with the measured frequency, 2.22778335(4) GHz, in the absence of any optical potentials. At this lattice depth, the slope for the differential light shift is 120 Hz/ E_r . The total variation of the light shift across the sample is less than 500 Hz, as estimated from the Ramsey decay time in absence of spin echo.

This residual light shift limits the coherence time of our standard Ramsey measurement to ~ 1 ms (triangle data points in Fig. 3c), and so we use a spin-echo pulse to mitigate the effects of single-particle dephasing. However, spatial variations of the differential light shift can in principle still influence the spin dynamics. Site-dependent shifts of the resonance frequency would appear in the Hamiltonian as an inhomogeneous ‘magnetic field’ term, $\delta_S S_i^z$, which can suppress spin exchange. For the conditions used in our experiments, we estimate that the spatial variations of the differential light shift are small enough that spin exchange remains near-resonant. On the basis of the measured Ramsey coherence times and the details of our system (optical lattice beams and the molecular ensemble), we calculate a system-averaged nearest-neighbour bias ($|\langle \delta_i - \delta_{i+1} \rangle|/h$) of 6 Hz, which is well below the expected exchange coupling, $|J_\perp/2h| = 52$ Hz. For neighbours separated by $\sqrt{2}a$ and $2a$, the corresponding shifts are 9 and 13 Hz, respectively. To demonstrate that the oscillation frequency in the Ramsey fringe contrast does not sensitively depend on the optical intensity, we made Ramsey contrast measurements

(Supplementary Fig. 2) for values of the lattice depth between $20E_r$ and $40E_r$ and found very good agreement among them.

Multipulse sequence. It is straightforward to understand how the multipulse sequence works for the case of two particles. With two molecules initially prepared in $|\downarrow\downarrow\rangle$, an initial $(\pi/2)_y$ -pulse transfers them to

$$\frac{1}{\sqrt{2}}(|\downarrow\rangle + |\uparrow\rangle) \otimes \frac{1}{\sqrt{2}}(|\downarrow\rangle + |\uparrow\rangle) = \frac{1}{2}(|\downarrow\downarrow\rangle + |\uparrow\uparrow\rangle + |\downarrow\uparrow\rangle + |\uparrow\downarrow\rangle) \quad (6)$$

Because of the spin-exchange term, $|\downarrow\uparrow\rangle$ and $|\uparrow\downarrow\rangle$ are not eigenstates of the Hamiltonian in equation (1). However, the three triplet states $|\downarrow\downarrow\rangle$, $|\uparrow\uparrow\rangle$ and $(|\downarrow\uparrow\rangle + |\uparrow\downarrow\rangle)/\sqrt{2}$ are eigenstates of the Hamiltonian, with eigenenergies 0, 0 and $J_\perp/2$, respectively. We note that a single $(\pi/2)_x$ -pulse can swap the states $|\downarrow\downarrow\rangle + |\uparrow\uparrow\rangle$ and $|\downarrow\uparrow\rangle + |\uparrow\downarrow\rangle$, and can thus act as an effective spin echo for these contributions to the two-particle wavefunction.

During the first free-evolution time, of duration $T/8$, $|\downarrow\downarrow\rangle$ and $|\uparrow\uparrow\rangle$ accumulate no phase, whereas $(|\downarrow\uparrow\rangle + |\uparrow\downarrow\rangle)/\sqrt{2}$ accumulates a phase $e^{-i(J_\perp/h)T/16}$. At this point the state is entangled. We then apply a $(-\pi/2)_x$ -pulse to swap the contributions from $|\downarrow\uparrow\rangle + |\uparrow\downarrow\rangle$ and $|\downarrow\downarrow\rangle + |\uparrow\uparrow\rangle$. This can alternatively be viewed as swapping the accrued phases. After another free-evolution time, of $T/4$, the $(\pi/2)_x$ -pulse swaps the phases again. This state then freely evolves for another time $T/8$, after which both $|\downarrow\uparrow\rangle + |\uparrow\downarrow\rangle$ and $|\downarrow\downarrow\rangle + |\uparrow\uparrow\rangle$ have accumulated the same phase, $e^{-i(J_\perp/h)T/8}$, and the state is, as a result, no longer entangled. In this way, the dephasing due to pairwise dipole–dipole interactions is cancelled. The centre $(\pi)_x$ -pulse and another pair of $(-\pi/2)_x$ - and $(\pi/2)_x$ -pulses are necessary for removing the single-particle inhomogeneity in addition to rephasing the dipole–dipole interactions. The effects of dipole–dipole interactions beyond that of isolated pairs of molecules are not removed by this particular multipulse sequence.

Interspecies two-body loss coefficient. To determine the appropriate two-body loss coefficient to describe the recombination of molecules in distinguishable rotational states, we measured losses for an incoherent mixture of $|0, 0\rangle$ and $|1, -1\rangle$ molecules in an optical dipole trap, as presented in Supplementary Fig. 3. We find a two-body loss coefficient of $9.0(4) \times 10^{-10} \text{ cm}^3 \text{ s}^{-1}$ for the mixture of rotational states that support resonant dipolar interactions³¹. This exceeds the value for molecules in different nuclear (hyperfine) states² (with no resonant dipolar coupling) by a factor of ~ 5 .

Theoretical modelling of the spin dynamics. Theoretical modelling of the spin dynamics observed with Ramsey spectroscopy shows similar oscillations and coherence times as our measurements, and the comparison can be used to estimate a filling factor of 5–10% for 10^4 molecules. Although exactly treating the many-body dynamics is intractable, at sufficiently small filling a ‘cluster expansion’ can be quite accurate. Here we separate N molecules into clusters, each containing at most 10 molecules, and solve exactly the spin dynamics within these clusters, neglecting intercluster interactions^{32,33}.

Results on the oscillation frequency, amplitude and decay time for the Ramsey contrast produced by the cluster expansion are all consistent with the experimental observations. We find that the cluster expansion depends relatively weakly on the assumed spatial distribution of the molecules (which we know only roughly), and that the main dependence of the contrast decay comes from the assumed filling, such that the filling estimate stated in the main text is reliable in this regard. A discussion of the convergence of the cluster expansion is far beyond the scope of the present work, but its uncertainties are the main source of the uncertainty in the estimated filling fraction.

The cluster expansion also indicates that the Ramsey contrast decay arises from many-body interaction effects other than either nearest-neighbour interactions or summing the dynamics of pairs of molecules. Although summing pairwise dynamics leads to a contrast decay time, τ , that decreases with increasing filling fraction, this τ is significantly larger than the cluster expansion results and requires unreasonably high lattice filling to match experimental results.

- Quémener, G., Bohn, J. L., Petrov, A. & Kotochigova, S. Universalities in ultracold reactions of alkali-metal polar molecules. *Phys. Rev. A* **84**, 062703 (2011).
- Witzel, W. M., de Sousa, R. & Das Sarma, S. Quantum theory of spectral-diffusion-induced electron spin decoherence. *Phys. Rev. B* **72**, 161306 (2005).
- Maze, J. R., Taylor, J. M. & Lukin, M. D. Electron spin decoherence of single nitrogen-vacancy defects in diamond. *Phys. Rev. B* **78**, 094303 (2008).

Carbon nanotube computer

Max M. Shulaker¹, Gage Hills², Nishant Patil³, Hai Wei⁴, Hong-Yu Chen⁵, H.-S. Philip Wong⁶ & Subhasish Mitra⁷

The miniaturization of electronic devices has been the principal driving force behind the semiconductor industry, and has brought about major improvements in computational power and energy efficiency. Although advances with silicon-based electronics continue to be made, alternative technologies are being explored. Digital circuits based on transistors fabricated from carbon nanotubes (CNTs) have the potential to outperform silicon by improving the energy-delay product, a metric of energy efficiency, by more than an order of magnitude. Hence, CNTs are an exciting complement to existing semiconductor technologies^{1,2}. Owing to substantial fundamental imperfections inherent in CNTs, however, only very basic circuit blocks have been demonstrated. Here we show how these imperfections can be overcome, and demonstrate the first computer built entirely using CNT-based transistors. The CNT computer runs an operating system that is capable of multitasking; as a demonstration, we perform counting and integer-sorting simultaneously. In addition, we implement 20 different instructions from the commercial MIPS instruction set to demonstrate the generality of our CNT computer. This experimental demonstration is the most complex carbon-based electronic system yet realized. It is a considerable advance because CNTs are prominent among a variety of emerging technologies that are being considered for the next generation of highly energy-efficient electronic systems^{3,4}.

CNTs are hollow, cylindrical nanostructures composed of a single sheet of carbon atoms, and have exceptional electrical, physical and thermal properties^{5–7}. They can be used to fabricate CNT field-effect transistors (CNFETs), which are promising candidate building blocks for the next generation of highly energy-efficient electronics^{1,2,8}; CNFET-based digital systems are predicted to be able to outperform silicon-based complementary metal-oxide-semiconductor (CMOS) technologies by more than an order of magnitude in terms of energy-delay product, a measure of energy efficiency^{2–4}.

Since the initial discovery of CNTs, there have been several major milestones for CNT technologies⁹: CNFETs, basic circuit elements (logic gates), a five-stage ring oscillator fabricated along a single CNT, a percolation-transport-based decoder, stand-alone circuit elements such as half-adder sum generators and D-latches, and a capacitive sensor interface circuit^{10–16}. Yet there remains a serious gap between these circuit demonstrations for this emerging technology and the first computers built using silicon transistors, such as the Intel 4004 and the VAX-11 (1970s). These silicon-based computers were fundamentally different from the above-mentioned CNFET-based circuits in several key ways: they ran stored programs, they were programmable (meaning that they could execute a variety of computational tasks through proper sequencing of instructions without modifying the underlying hardware¹⁷) and they implemented synchronous digital systems incorporating combinational logic circuits interfaced with sequential elements such as latches and flip-flops¹⁸.

It is well known that substantial imperfections inherent in CNT technology are the main obstacles to the demonstration of robust and complex CNFET circuits¹⁹. These include mis-positioned and metallic CNTs. Mis-positioned CNTs create stray conducting paths leading

to incorrect logic functionality, whereas metallic CNTs have little or no bandgap, resulting in high leakage currents and incorrect logic functionality²⁰. The imperfection-immune design methodology, which combines circuit design techniques with CNT processing solutions, overcomes these problems^{20,21}. It enables us to demonstrate, for the first time, a complete CNT computer, realized entirely using CNFETs. Similar to the first silicon-based computers, our CNT computer, which is a synchronous digital system built entirely from CNFETs, runs stored programs and is programmable. Our CNT computer runs a basic operating system that performs multitasking, meaning that it can execute multiple programs concurrently (in an interleaved fashion). We demonstrate our CNT computer by concurrently executing a counting program and an integer-sorting program (coordinated by a basic multitasking operating system), and also by executing 20 different instructions from the commercial MIPS instruction set²².

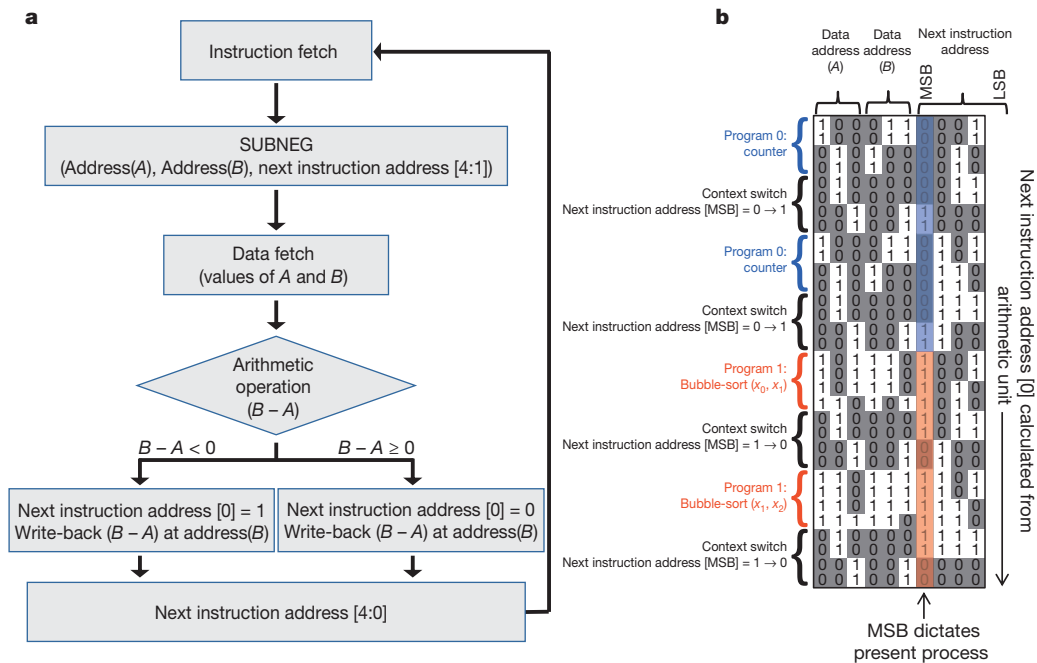
The CNT computer is a one-instruction-set computer, implementing the SUBNEG (subtract and branch if negative) instruction, inspired by early work in ref. 23. We implement the SUBNEG instruction because it is Turing complete and thus can be used to re-encode and perform any arbitrary instruction from any instruction-set architecture, albeit at the expense of execution time and memory space^{24,25}. The SUBNEG instruction is composed of three operands: two data addresses and a third partial next instruction address (the CNT computer itself completes the next instruction address, allowing for branching to different instruction addresses). The SUBNEG instruction subtracts the value of the data stored in the first data address from the value of the data stored in the second data address, and writes the result at the location of the second data address.

The next instruction address is calculated to be one of two possible branch locations, depending on whether the result of the subtraction is negative. The partial next instruction address given by the present SUBNEG instruction omits the least significant bit. The least significant bit is calculated by the CNT computer, on the basis of whether the result of the SUBNEG subtraction was negative. This bit, concatenated with the partial next instruction address given in the SUBNEG instruction, makes up the entire next instruction address. A diagram showing the SUBNEG implementation is shown in Fig. 1a.

As our operating system, we implement non-pre-emptive multitasking, whereby each program performs a self-interrupt and voluntarily gives control to another task²⁶. To perform this context switch, the instruction memory is structured in blocks, and each block contains a different program. To perform the self-interrupt, the running program stores a next instruction address belonging to a different program block; thus, the other program begins execution at this time. During the context switch, the CNT computer updates a process ID bit in memory, which indicates the program running at present. An example of the operating system running two different programs concurrently is shown in Fig. 1b.

The circuitry of the CNT computer is entirely composed of CNFETs, and the instruction and data memories are implemented off-chip, following the von Neumann architecture and the convention of most computers today. The off-chip memories perform no operation other

¹Stanford University, Gates Building, Room 331, 353 Serra Mall, Stanford, California 94305, USA. ²Stanford University, Gates Building, Room 358, 353 Serra Mall, Stanford, California 94305, USA. ³SK Hynix Memory Solutions, 3103 North First Street, San Jose, California 95134, USA. ⁴Stanford University, Gates Building, Room 239, 353 Serra Mall, Stanford, California 94305, USA. ⁵Stanford University, Paul G. Allen Building, Room B113X, 420 Via Ortega, Stanford, California 94305, USA. ⁶Stanford University, Paul G. Allen Building, Room 312X, 420 Via Ortega, Stanford, California 94305, USA. ⁷Stanford University, Gates Building, Room 334, 353 Serra Mall, Stanford, California 94305, USA.



than performing a single read or a single write in a clock cycle. The address, data (for write), and read and write enable signals are provided by the CNT computer; the values, once read, are stored in D-latches in the CNT computer, built entirely using CNFETs. A full schematic of the CNT computer is shown in Fig. 2a. The CNT computer performs four tasks.

(1) Instruction fetch: this task supplies instruction memory with the address to read. On the first clock (Clock1), the SUBNEG instruction is read from the instruction memory and saved in a bank of ten D-latches. The SUBNEG instruction contains the partial next instruction address

(as explained above), and the addresses of the two single-bit data values to operate on (represented as [A] and [B], both of which comprise three bits).

(2) Data fetch: this task supplies the data memory with the addresses given by the SUBNEG instruction to read. On Clock1, the first data address ([A]) is read and the value is saved in a D-latch. On the second clock (Clock2), the second data address ([B]) is read and the value is saved in another D-latch.

(3) Arithmetic operation: this task performs the computation (subtraction and comparison with zero) on the two data values supplied by the data-fetch unit.

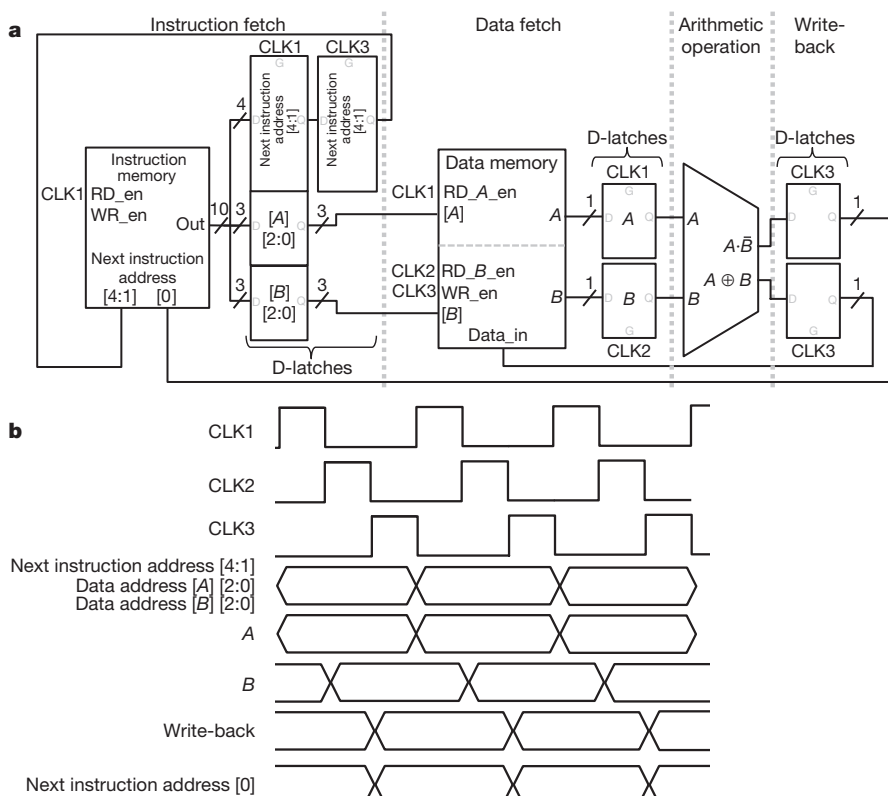


Figure 2 | Schematic of CNT computer.

a, Schematic of the entire CNT computer, composed of the four subunits: instruction fetch, data fetch, arithmetic operation and write-back. All components apart from the memory are implemented entirely using CNFETs. CLK1–CLK3, Clock1–Clock3; D, D-latch input; Q, D-latch output; G, D-latch clock; RD_en, read enable (instruction memory); WR_en, write enable (instruction memory); RD_A_en, read enable address A (data memory); RD_B_en, read enable address B (data memory); Data_in, data for data memory write. **b**, Timing diagram of the CNT computer. The lines show the waveforms corresponding to each signal; of particular note are the transitions of the lower five signals with respect to the clock signals.

(4) Write-back: this task writes back the result of the SUBNEG ($B - A$) in the data memory at the address of the second data address. On the third clock (Clock3), the result from the arithmetic-operation unit is saved in two D-latches. Simultaneously, Clock3 enables the write-back to the data memory. D-latches from the instruction-fetch unit supply the data address, and the D-latch from the write-back stage supplies the value to be written.

A timing diagram depicting the above description and using three non-overlapping clocks is shown in Fig. 2b.

The CNFET computer is composed of 178 CNFETs, with each CNFET comprising ~ 10 –200 CNTs, depending on relative sizing of the widths of the CNFETs. Figure 3 shows transistor-level schematics of the sub-components, D-latches and the arithmetic unit. We use logic circuits that use only p-type transistors, because our CNFETs are p-type without modifications. Consequently, relative sizing of the widths of pull-up and pull-down CNFETs is crucial; the ratio of all pull-up CNFET widths to pull-down CNFET widths in our design is either 20:1 or 10:1 (Methods). There is a maximum of seven stages of cascaded logic in the computer, demonstrating our ability to cascade combinational logic stages, which is a necessity in realizing large digital systems.

The CNT-specific fabrication process is based on the process described in refs 21, 23, 27, and is described in detail in Methods. Importantly, the fabrication process is completely silicon-CMOS compatible owing to its low thermal budget (125°C). We use standard cells for our sub-systems, designed following the imperfection-immune methodology, which renders our circuits immune to both mis-positioned and metallic CNTs. Because this method ensures that the immunity to CNT imperfections is encapsulated entirely within standard cells, the fabrication is

completely insensitive to the exact positioning of CNTs on the wafer and there is no per-unit customization, rendering our processing and design VLSI (very large-scale integration) compatible. The entire CNT computer is fabricated completely within a die on a single wafer. Each die contains five CNT computers, and each wafer contains 197 dies. There is no customization of any sort after circuit fabrication: all of the CNFETs and interconnects are predetermined during design, and there is no post-fabrication selection, configuration or fine-tuning of functional CNFETs. Just like any von Neumann computer, off-chip interconnects are used for connections to external memories. Our CNT-specific fabrication process and imperfection-immune design enables high yield and robust devices; waveforms of 240 subsystems (40 arithmetic logic units and 200 D-latches) from across a wafer are shown in Fig. 3. The yield of the subsystems, such as D-latches, typically ranges from 80% to 90%. The primary causes of yield loss—particles resulting in broken lithography patterns, adhesion issues with metal lift-off and variations in machine etch rates—are consequences of the limitations of performing all fabrication steps in-house in an academic fabrication facility.

A SEM image of a fabricated CNT computer is shown in Fig. 4a. To demonstrate the working CNT computer, we perform multitasking with our basic operating system, concurrently running a counter program and an integer-sorting program (performing the bubble-sort algorithm). Although CNFET circuits promise improved speed^{2,4,8}, our computer runs at 1 kHz. This is not due to the limitations of the CNT technology or our design methodology, but instead is caused by capacitive loading introduced by the measurement setup, the $1\text{-}\mu\text{m}$ minimum lithographic feature size possible in our academic fabrication facility, and CNT density and contact resistance (Methods). The measured and expected

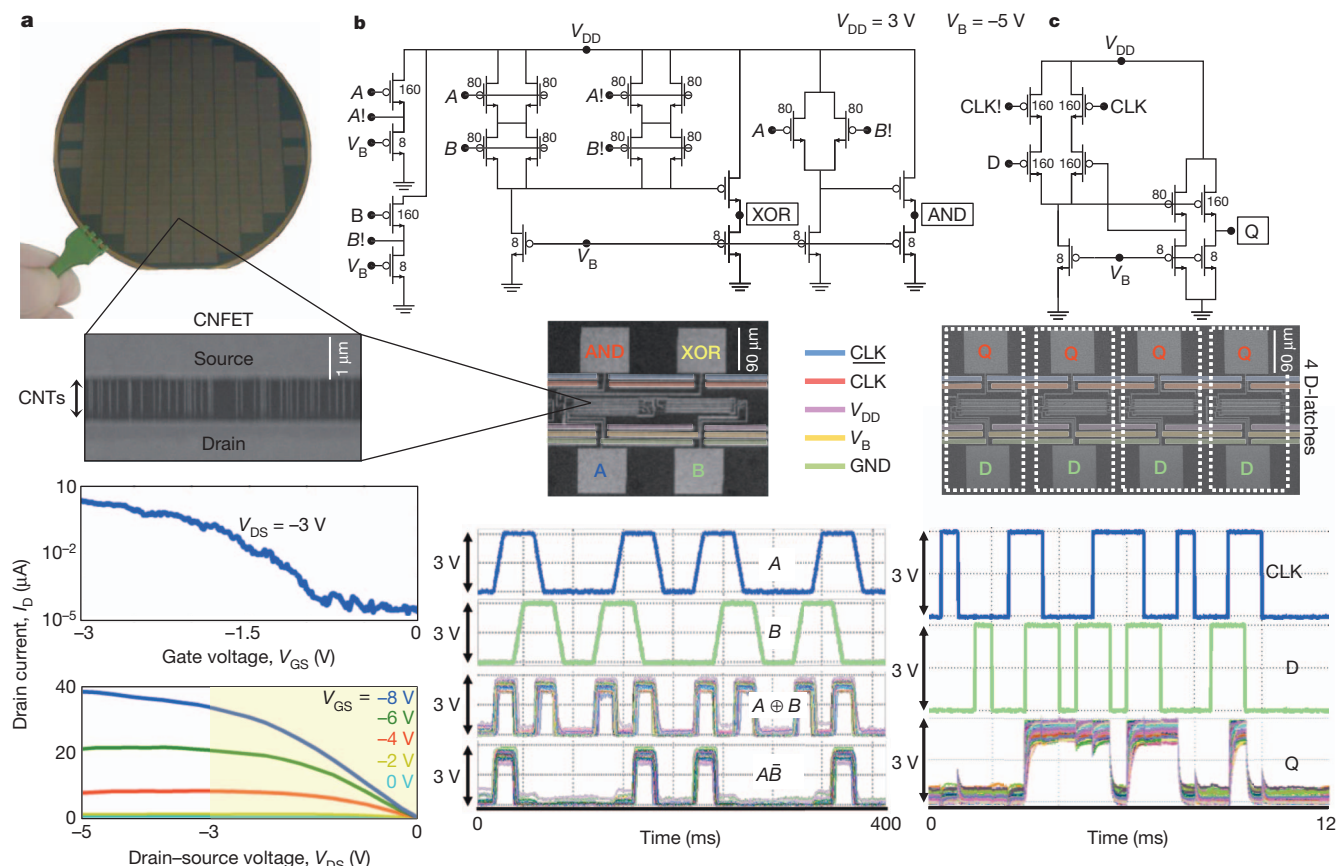


Figure 3 | Characterization of CNFET subcomponents. **a**, Top: Final 4-inch wafer after all fabrication. Middle: scanning electron microscope (SEM) image of a CNFET, showing source, drain and CNTs extending into the channel region. Bottom: Measured characterization (current–voltage) curves of a typical CNFET. The yellow highlighted region of the I_D – V_{DS} curve shows the biasing region that the CNFET operates in for the CNT computer. **b**, Top:

transistor-level schematic of arithmetic unit. Numbers are width of transistors (in micrometres). Middle: SEM of an arithmetic unit. Bottom: measured outputs from 40 different arithmetic units, all overlaid. **c**, Top: transistor-level schematic of D-latches. Numbers are width of transistors (in micrometres). Middle: SEM of a bank of 4 D-latches. Bottom: measured outputs from 200 different D-latches, all overlaid.

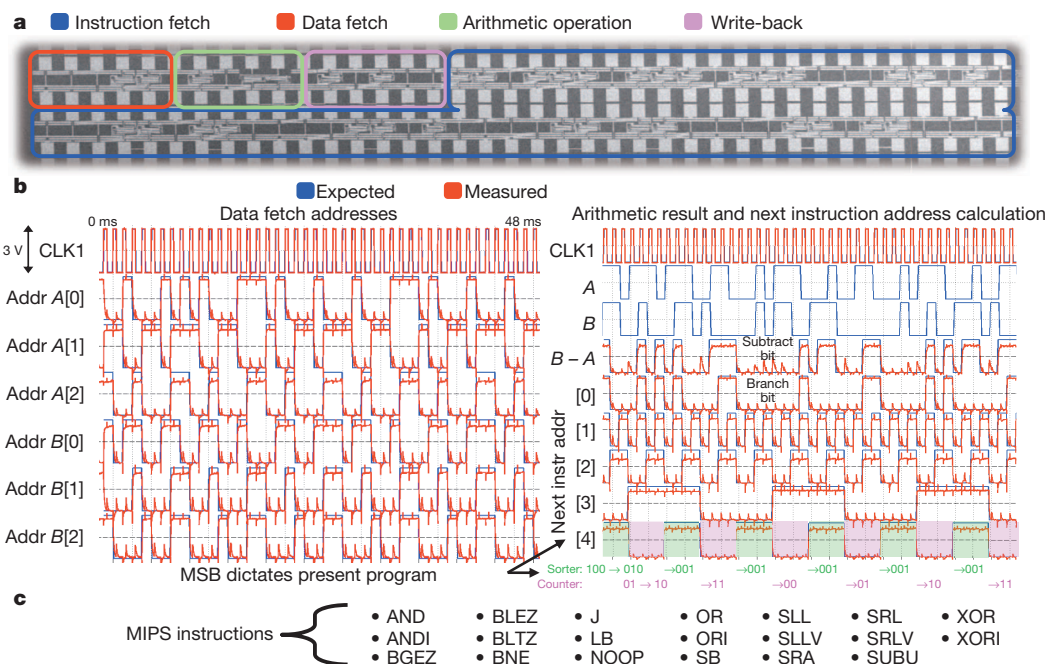


Figure 4 | CNT computer results. **a**, SEM of an entire CNT computer. **b**, Measured and expected output waveforms for a CNT computer, running the program shown in Fig. 1b. The exact match in logic value of the measured and expected output shows correct operation. As shown by the MSB (denoted [4]) of the next instruction address, the computer is switching between

performing counting and sorting (bubble-sort algorithm). The running results of the counting and sorting are shown in the rows beneath the MSB of the next instruction address. **c**, A list of the 20 MIPS instructions tested on the CNT computer.

outputs from the CNT computer (Fig. 4b) show correct operation. To demonstrate the flexibility and ability of the SUBNEG computer to implement any arbitrary instruction, we additionally perform 20 MIPS instructions (Fig. 4c) on the CNT computer. Although the CNT computer operates on single-bit data values, this is not a fundamental limitation, because any multibit computation can be performed with a single-bit computer through serial computation²³. Additionally, having shown the ability to cascade logic, fabricating a larger multibit CNT computer is not a fundamental obstacle, but rather affects only yield; as a demonstration, we show a two-bit arithmetic logic unit (composed of 96 CNFETs with a maximum of 15 stages of cascaded logic) in Extended Data Fig. 2 (see also Methods).

We have reported a CNT computer fabricated entirely from CNFETs, and have demonstrated its ability to run programs, to run a basic operating system that performs multitasking, and to execute MIPS instructions. To achieve this we used the imperfection-immune design methodology and developed robust and repeatable CNT-specific design and processing. This demonstration confirms that CNFET-based circuits are a feasible and plausible emerging technology.

METHODS SUMMARY

The fabrication process is depicted in Extended Data Fig. 1. The CNTs are grown on a quartz substrate to yield highly aligned CNTs¹⁴, and are transferred onto the target SiO₂ wafer¹⁴. Before CNT transfer, the wafer undergoes processing to define bottom-layer wires and the local back gates of the transistors²⁸. Lithographically defined trenches are etched using a combination of dry plasma etch followed by wet etch, and are filled by electron-beam evaporation of platinum and smoothed by a subsequent plasma sputter etch. A 24-nm high-*k* dielectric of Al₂O₃ is deposited by atomic-layer deposition, and contact holes are etched through this layer to the embedded metal wires and gates through another combined dry- and wet-etch process. After CNT transfer, the source and drain (bilayers of palladium and platinum) are lithographically defined through a lift-off process, and mis-positioned CNTs are etched away using optical lithography followed by oxygen plasma¹⁵. A metal layer of gold is lithographically patterned with lift-off and connects every other source and drain, and separately connects every gate, effectively forming a single CNFET composed of all of the single CNFETs in parallel. Electrical breakdown is performed once on this entire structure to remove >99.99% of metallic

CNTs^{29,30}. This gold layer is then selectively etched away, and the top metal layer connecting the circuit in the proper configuration is lithographically patterned and deposited with lift-off.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 May; accepted 24 July 2013.

- Franklin, A. D. *et al.* Sub-10 nm carbon nanotube transistor. *Nano Lett.* **12**, 758–762 (2012).
- Wei, L., Frank, D., Chang, L. & Wong, H.-S. P. in *Proc. 2009 IEEE Intl Electron Devices Meeting* 917–920 (IEEE, 2009).
- Chang, L. in *Short Course IEEE Intl Electron Devices Meeting* (IEEE, 2012).
- Nikonov, D. & Young, I. in *Proc. 2012 IEEE Intl Electron Devices Meeting* 24–25 (IEEE, 2012).
- Javey, A., Guo, J., Wang, Q., Lundstrom, M. & Dai, H. Ballistic carbon nanotube transistors. *Nature* **424**, 654–657 (2003).
- Javey, A., Wang, Q., Kim, W. & Dai, H. in *2003 Intl Electron Devices Meeting Tech. Digest* 31–32 (IEEE, 2003).
- Appenzeller, J. Carbon nanotubes for high-performance electronics—progress and prospect. *Proc. IEEE* **96**, 201–211 (2008).
- Deng, J. *et al.* in *Proc. 2007 IEEE Intl Solid State Circuits Conf.* 70–78 (IEEE, 2007).
- Iijima, S. Helical microtubules of graphitic carbon. *Nature* **354**, 56–58 (1991).
- Martel, R. A., Schmidt, T., Shea, H. R., Hertel, T. & Avouris, P. Single- and multi-wall carbon nanotube field-effect transistors. *Appl. Phys. Lett.* **73**, 2447 (1998).
- Tans, S. J., Verschueren, A. R. & Dekker, C. Room-temperature transistor based on a single carbon nanotube. *Nature* **393**, 49–52 (1998).
- Chen, Z. *et al.* An integrated logic circuit assembled on a single carbon nanotube. *Science* **311**, 1735 (2006).
- Cao, Q. *et al.* Medium-scale carbon nanotube thin-film integrated circuits on flexible plastic substrates. *Nature* **454**, 495–500 (2008).
- Patil, N., Lin, A., Myers, E. R., Wong, H.-S. P. & Mitra, S. in *Proc. Symp. VLSI Tech.* 205–206 (2008).
- Patil, N. *et al.* Scalable carbon nanotube computational and storage circuits immune to metallic and mis-positioned carbon nanotubes. *IEEE Trans. NanoTechnol.* **10**, 744–750 (2011).
- Shulaker, M. *et al.* in *Proc. 2013 IEEE Intl Solid State Circuits Conf.* 112–113 (IEEE, 2013).
- von Neumann, J. First draft of a report on the EDVAC. *Ann. Hist. Comput.* **15**, 27–75 (1993).
- McCluskey, E. J. *Logic Design Principles with Emphasis on Testable Semicustom Circuits* (Prentice-Hall, 1986).
- Cao, Q. *et al.* Arrays of single-walled carbon nanotubes with full surface coverage for high-performance electronics. *Nature Nanotechnol.* **8**, 180–186 (2013).

20. Zhang, J. *et al.* Robust digital VLSI using carbon nanotubes. *IEEE Trans. CAD* **31**, 453–471 (2012).
21. Patil, N. *Design and Fabrication of Imperfection-Immune Carbon Nanotube Digital VLSI Circuits*. PhD thesis, Stanford Univ. (2010).
22. Patterson, D. A. & Hennessy, J. L. *Computer Architecture* (Kaufmann, 1990).
23. Lin, A. *Carbon Nanotube Synthesis, Device Fabrication, and Circuit Design for Digital Logic Applications*. PhD thesis, Stanford Univ. (2010).
24. Herken, R. (ed.) *The Universal Turing Machine: A Half-Century Survey* (Springer, 1995).
25. Nürnberg, P., Uffe, W. & Hicks, D. A grand unified theory for structural computing. *Metainformatics* **3002**, 1–16 (2004).
26. Jeffay, K., Donald, S. F. & Martel, C. U. in *Proc. Real-Time Systems Symposium* 129–139 (IEEE, 1991).
27. Shulaker, M. *et al.* Linear increases in carbon nanotube density through multiple transfer technique. *Nano Lett.* **11**, 1881–1886 (2011).
28. Bachtold, A., Hadley, P., Nakanishi, T. & Dekker, C. Logic circuits with carbon nanotube transistors. *Science* **294**, 1317–1320 (2001).
29. Collins, P. G., Arnold, M. S. & Avouris, P. Engineering carbon nanotubes and nanotube circuits using electrical breakdown. *Science* **292**, 706–709 (2001).
30. Patil, N. *et al.* in *Proc. 2009 IEEE Intl Electron Devices Meeting* 573–576 (IEEE, 2009).

Acknowledgements We acknowledge the support of the NSF (CISE) (CNS-1059020, CCF-0726791, CCF-0702343, CCF-0643319), FCRP C2S2, FCRP FENA, STARNet SONIC and the Stanford Graduate Fellowship and the Hertz Foundation Fellowship (M.M.S.). We also acknowledge Z. Bao, A. Lin, H. (D.) Lin, M. Rosenblum, and J. Zhang for their advice and collaborations.

Author Contributions M.M.S. led and was involved in all aspects of the project, did all of the fabrication and layout designs, and contributed to the design and testing. G.H. wrote the SUBNEG and testing programs, and contributed to the design and testing. N.P. contributed to the design, and N.P., H.W. and H.-Y.C. contributed to developing fabrication processes. H.-S.P.W. and S.M. were in charge and advised on all parts of the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M.S. (maxms@stanford.edu).

METHODS

The fabrication process is depicted in Extended Data Fig. 1.

CNT growth and transfer. The CNTs are grown by chemical-vapour deposition with methane at 865 °C. The growth substrate is an annealed quartz substrate, with parallel catalyst stripes of iron lithographically patterned on the wafer. Quartz is used to achieve 99.5% alignment of the CNTs, which align along the crystalline boundary owing to a minimized Lennard-Jones potential in this orientation¹⁴. After growth, the quartz wafer with CNTs is coated with 150 nm gold, and a thermal release tape is applied on top of the gold. When this tape is peeled from the wafer, it peels off the gold with embedded CNTs from the quartz wafer. The tape is then applied onto the target wafer and heated to 125 °C, at which point the thermal release tape loses adhesion and is removed from the wafer, leaving the gold with embedded CNTs on the target wafer. The surface of the wafer undergoes oxygen and argon plasma etching to remove any residue from the tape, followed by a selective wet etch to remove the gold, leaving exposed, highly aligned CNTs on the wafer¹⁴.

Local back gate. Before transfer, the target wafer is first prepared, starting with a silicon wafer with 110 nm thermal oxide growth. To form the local back gate²⁸ and bottom layer of wires, a two-layer resist stack is lithographically patterned on the surface. Following development of the pattern, the wafer goes through a quick oxygen plasma de-scum, followed by an anisotropic O₂/SF₆ plasma etch. After the plasma etch, a quick HF dip is used to smooth the surface and remove any side-wall deposition from the plasma etching. Next, an adhesion layer of Ti followed by Pt is evaporated, filling the trenches etched in the previous step. The bilayer of resist is dissolved away, lifting off the extra metal and leaving the metal in the trenches. An argon sputter etch follows, and, owing to the difference in etch rate between the Pt and SiO₂, the surface of the wafer is smoothed until the offset between the local back gate height and the wafer is less than a nanometre.

Initial transistor fabrication. We use ~24 nm Al₂O₃ as our high-*k* back-gate dielectric. This is deposited through atomic-layer deposition on the wafer described above, covering the local back gates and bottom-level wires. Before CNT transfer, the deposited surface undergoes an oxygen plasma etch to clean the surface of any contaminants and a forming gas anneal, followed by the CNT transfer process described above. Immediately following transfer is source-drain definition of the individual transistors. A bilayer of resist is patterned and developed, and a bilayer of 20 nm Pd and 20 nm Pt is deposited for both the source and drains. This is followed by a traditional lift-off process. In addition to the source and drain, a second layer of metal wiring is patterned and deposited. This second layer of metal wiring is permanent through the rest of the process. After the metal deposition, mis-positioned and unneeded CNTs are removed by covering the active area of the transistors with photoresist and etching away the unprotected CNTs with oxygen plasma. The layout of the active area of the transistors follows the mis-positioned CNT immune design^{20,21}, and guarantees that no mis-positioned CNTs can cause incorrect logic function. This renders the circuit immune to mis-positioned CNTs. Contacts to the bottom-layer wires and local back gates are lithographically defined and etched with an Ar/CL₂/BCL₃ plasma etch, followed by HF dip, with the embedded metal acting as a natural etch stop.

Metallic CNT removal. To ensure high $I_{\text{on}}/I_{\text{off}}$ ratios and correct logic functionality, it is necessary to remove >99.99% of the metallic CNTs from the circuit, while leaving the semiconducting CNTs predominantly intact. This is achieved through electrical breakdown, which biases the gate of the transistor to turn the semiconducting CNTs off, and pulses a large current through the metallic CNTs, causing joule self-heating until the metallic CNTs oxidize and are removed, thus no longer conducting current²⁹. Rather than performing breakdown on the individual transistors, we employ VLSI-compatible metallic CNT removal³⁰ (VMR). VMR allows electrical breakdown to be performed on the chip scale. To do so, we lithographically define and pattern a gold layer through the lift-off processes described above. The gold is patterned to short every gate, source and drain together. This effectively forms a single large CNFET, composed of all of the single CNFETs connected in parallel. The shorted structures make use of the power rails and clock distribution networks to minimize area overhead. We then perform electrical breakdown on the entire structure once, enabling quick and efficient breakdown of hundreds of transistors and thousands of CNTs simultaneously (though this is not a fundamental limitation of the size of a VMR structure). After electrical

breakdown, the gold layer is removed. The third and final metal layer of Pt with an adhesion layer of Ti is deposited and lifted off, forming the final circuit layout configuration.

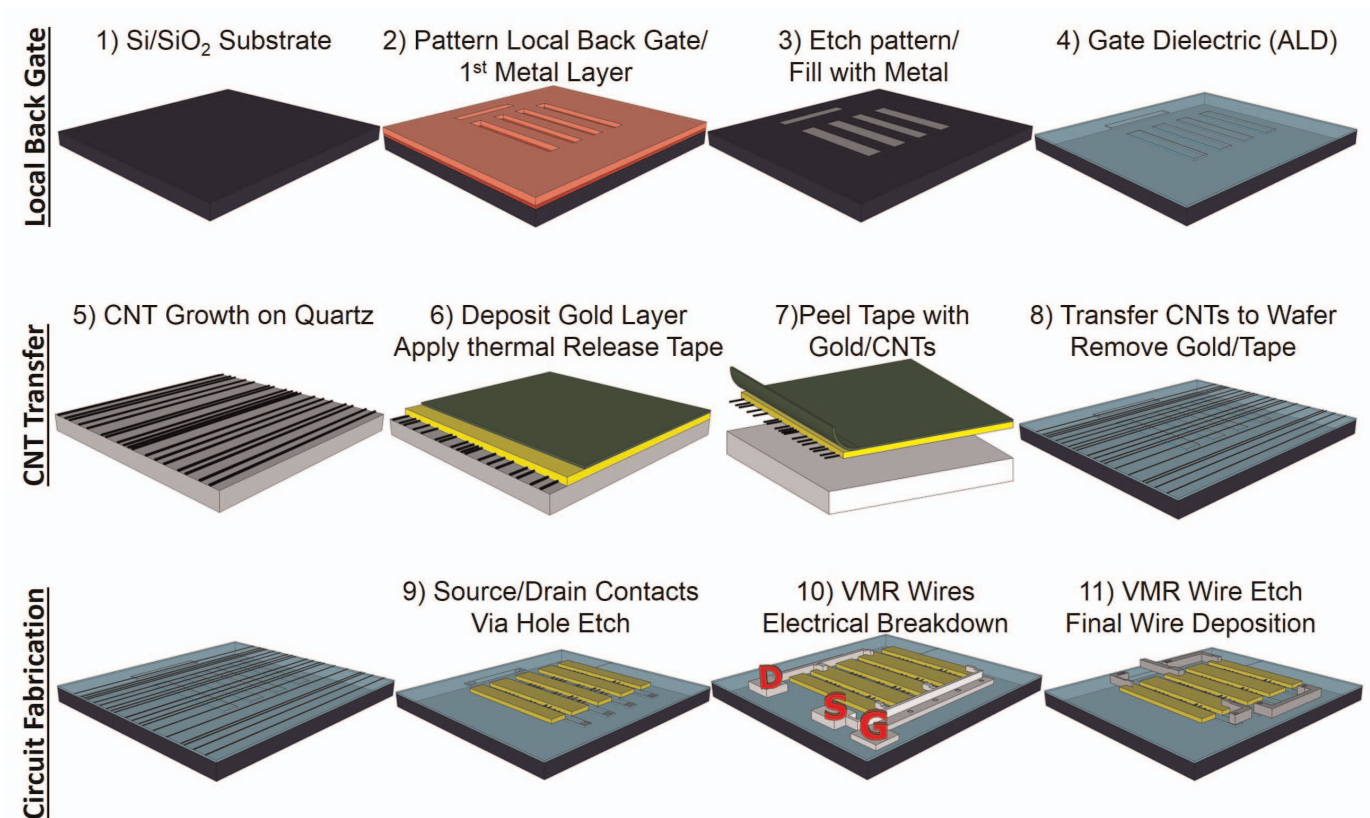
Test set-up. As shown in Fig. 4a, the CNT computer has four rows of probe pads, each containing 39 pads. A custom probe card is used to probe all of the pads simultaneously, although many of the pads are unused (and are simply present to ensure that the probe tips from the probe card always land on metal). Through the probe card, the pads are either connected to a supply voltage (V_{DD} , GND, V_{BIAS}) or to the inputs or outputs of the computer (the address outputs and input values to and from the off-chip memories). All other connections are made on-chip, as shown in Extended Data Fig. 3. A National Instrument DAQ (data acquisition hardware, #9264) is used to interface with the probe card and read and write the inputs and, respectively, outputs to the CNT computer, and Agilent oscilloscopes (#2014A) are additionally used to record the analogue traces of the outputs of the CNT computer (Fig. 4b).

Biasing. The biasing scheme for the circuits is shown in Fig. 3, with $V_{\text{DD}} = 3 \text{ V}$ and $V_{\text{BIAS}} = -5 \text{ V}$. There is no individual tuning of biasing voltages for individual transistors. Scaled supply voltages can be achieved by scaling the transistor channel lengths from 1 μm at present (due to the limitations of academic fabrication capabilities) to smaller channel lengths¹.

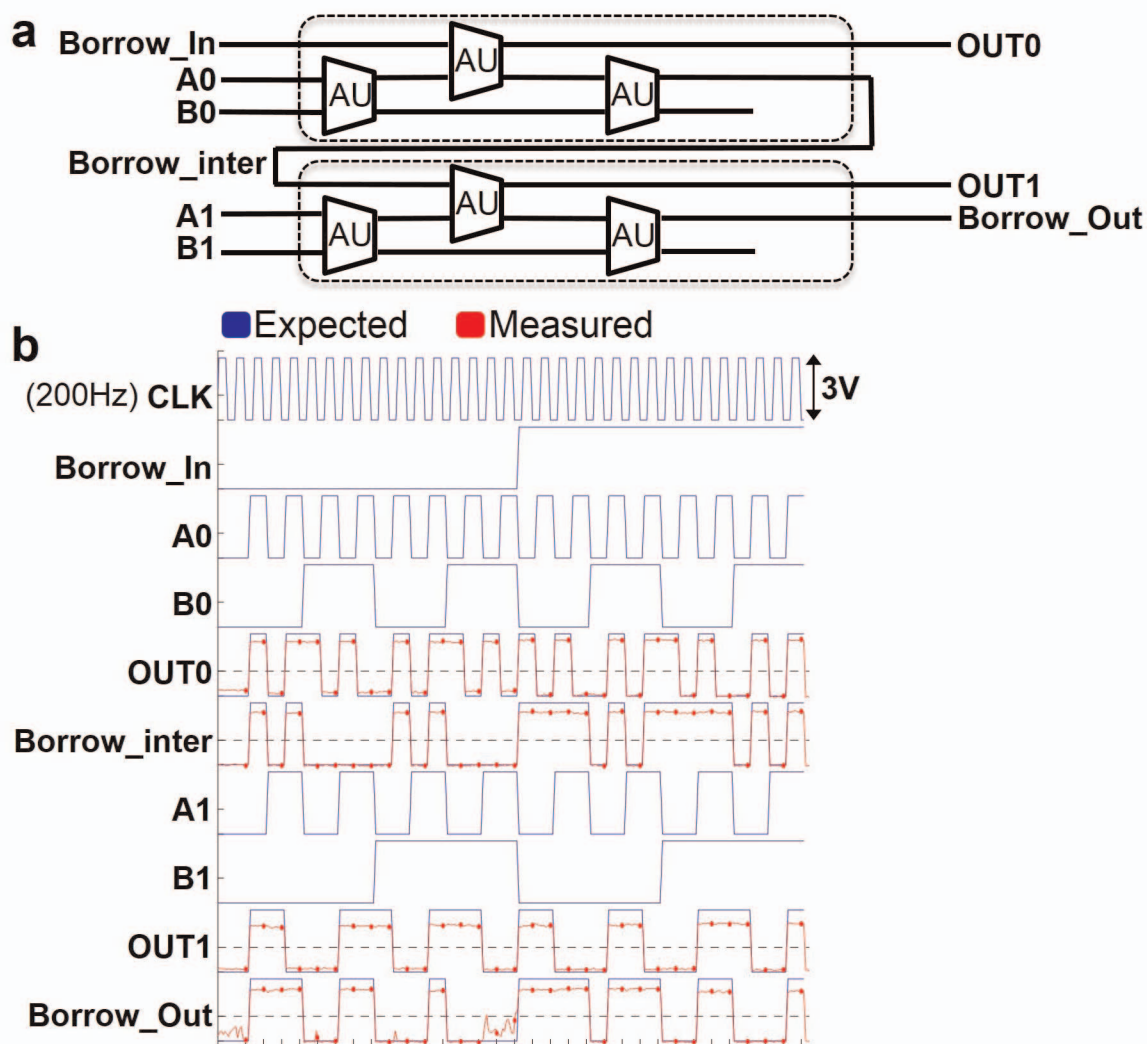
Speed. The probe pads and probe card with connecting wires used to connect to the CNT computer add additional capacitive loading to the circuit, limiting the frequency of operation to 1 kHz. However, this is not a fundamental limitation, because commercial chips are packaged and connected to memory and external devices without the use of probe cards, greatly reducing parasitic capacitances. The speed is also limited by the fact that the CNFET gate length is ~1 μm , set by the minimum lithographic feature that can be patterned in our academic clean-room; in field-effect transistors, on-current increases as the gate length decreases¹. Lithographic overlay accuracy of ~200 nm further increases parasitic capacitances resulting in reduced speed. Moreover, the CNT density in this work is ~5 CNTs per micrometre, whereas the target CNT density for increased current drive is 100–200 CNTs per micrometre⁸. Several published approaches show promising methods of achieving this target CNT density²⁷. CNT contact resistance must also be improved for high-performance circuits, and is another source of variation between devices.

PMOS-only logic. Logic circuits which use only p-type transistors are known as PMOS-only logic. The design of PMOS-only logic, which is well documented in the literature, is shown in Extended Data Fig. 4. Extended Data Fig. 4a depicts a PMOS-only inverter, whereas Extended Data Fig. 4b depicts a PMOS-only NAND gate. As is apparent from comparison of the two circuits, the pull-down network is always a single p-type transistor, whose gate is biased to remain on continuously. The pull-up network follows the design of typical CMOS circuits. The p-type transistors in the pull-up network create a conducting path from the output to V_{DD} when the output should be logic 1. When the output should be logic 0, the pull-up network is designed to no longer have a conducting path to V_{DD} , and, thus, the single p-type transistor in the pull-down network pulls the output to logic 0. The relative sizing of the pull-up network and pull-down network is critical, because the pull-down network is always biased on. Thus, when the pull-up network should pull the output to logic 1, the pull-down network will still be attempting to pull the output to logic 0. Thus, in our design, the transistors in the pull-up networks are always sized with a width of 10–20 times the pull-down transistor width. Exact transistor sizing is shown in Fig. 3.

Multibit arithmetic unit. Additionally, having shown the ability to cascade logic, fabricating a larger multibit CNT computer is not a fundamental obstacle, but rather only affects yield; as a demonstration, we show a two-bit arithmetic unit (composed of 96 CNFETs with a maximum of 15 stages of cascaded logic). The two-bit arithmetic unit is shown in Extended Data Fig. 2. The output waveform tests for all possible inputs, and shows correct operation. Additionally, we show that the circuits regenerate the signal between stages, a necessity for cascading digital logic, by highlighting the noise in the 'borrow out' output. Even with noise somewhere within the arithmetic unit (which can have multiple causes: a stage with low swing, electrical noise on the inputs, mobile charges in an oxide and so on), owing to the gain of each stage the final output levels (logic 0 and logic 1) always stay either below or above the threshold for logic 0 or logic 1, respectively (as shown by the horizontal black dotted line).

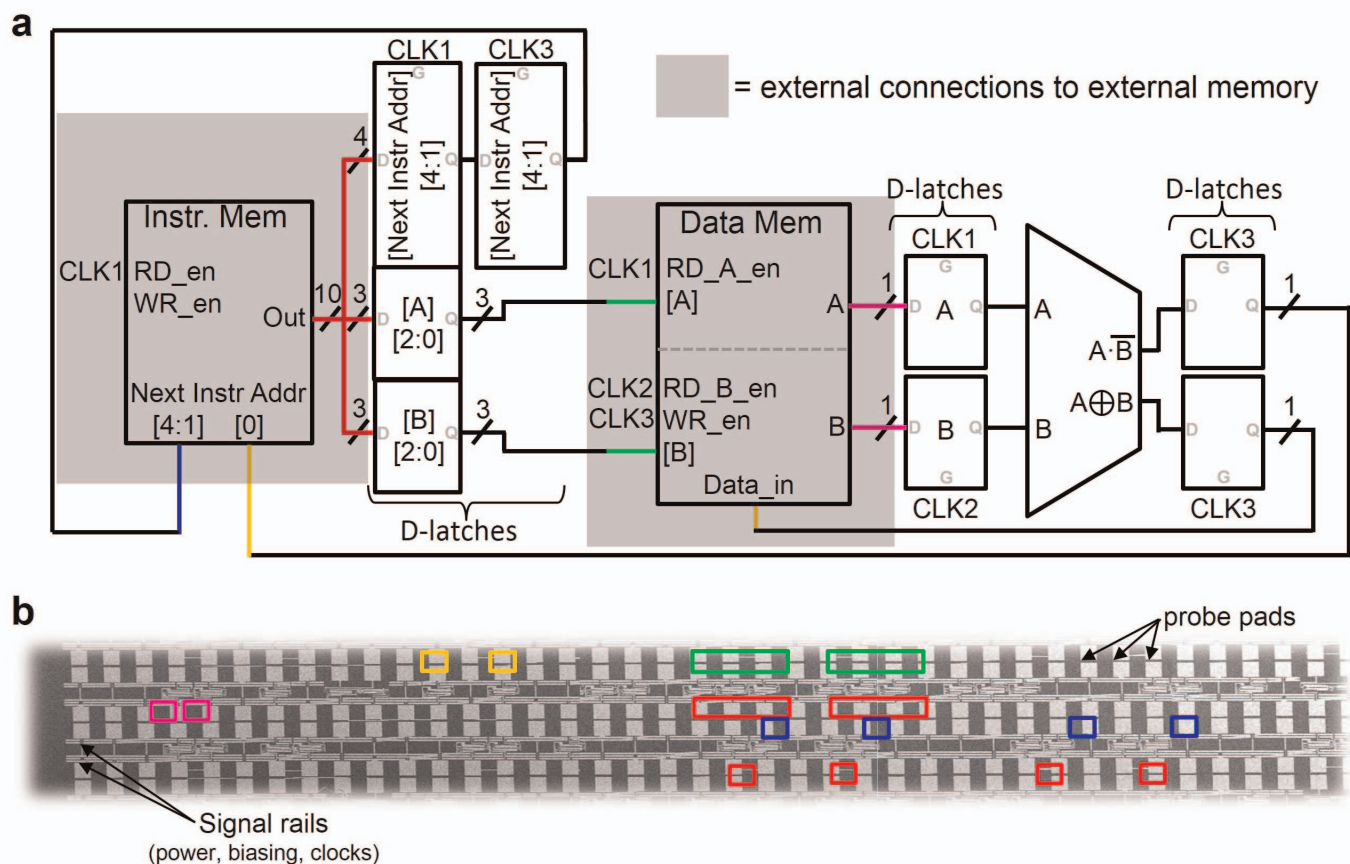


Extended Data Figure 1 | Fabrication flow for the CNT computer. Steps 1–4 prepare the final substrate for circuit fabrication. Steps 5–8 transfer the CNTs from the quartz wafer (where highly aligned CNTs are grown) to the final SiO₂ substrate. Steps 9–11 continue final device fabrication on the final substrate.



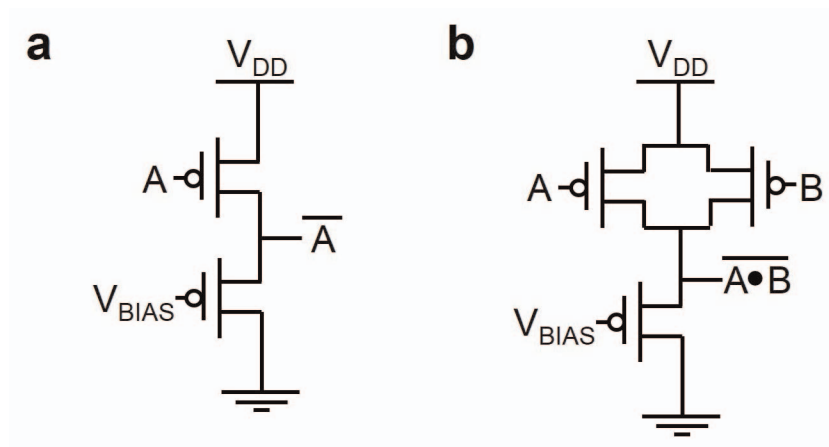
Extended Data Figure 2 | Multibit arithmetic unit. **a**, Schematic of a two-bit arithmetic unit, comprising six individual arithmetic logic units (ALU) as shown in Fig. 3b. **b**, Measured and expected output waveforms testing all

possible input combinations of the two-bit arithmetic unit, showing correct operation.



Extended Data Figure 3 | Internal versus external connections of CNT computer. **a**, Schematic of the CNT computer, showing that all connections are fabricated on-chip and that only signals reading or writing to or from an

external memory are connected off-chip. **b**, SEM of the CNT computer, showing which connections are made to and from the CNT computer from the probe pads. The SEM is colour-coded to match the coloured wires in **a**.



Extended Data Figure 4 | PMOS-only logic schematics. **a**, Schematic of PMOS-only inverter. **b**, Schematic of PMOS-only NAND gate.

Alkane desaturation by concerted double hydrogen atom transfer to benzyne

Dawen Niu¹, Patrick H. Willoughby¹, Brian P. Woods¹, Beeraiah Baire¹ & Thomas R. Hoye¹

The removal of two vicinal hydrogen atoms from an alkane to produce an alkene is a challenge for synthetic chemists^{1,2}. In nature, desaturases and acetylenases are adept at achieving this essential oxidative functionalization reaction, for example during the biosynthesis of unsaturated fatty acids³, eicosanoids, gibberellins⁴ and carotenoids⁵. Alkane-to-alkene conversion almost always involves one or more chemical intermediates in a multistep reaction pathway; these may be either isolable species (such as alcohols or alkyl halides) or reactive intermediates (such as carbocations, alkyl radicals, or σ -alkyl-metal species). Here we report a desaturation reaction of simple, unactivated alkanes that is mechanistically unique. We show that benzyne is capable of the concerted removal of two vicinal hydrogen atoms from a hydrocarbon. The discovery of this exothermic, net redox process was enabled by the simple thermal generation of reactive benzyne intermediates through the hexadehydro-Diels–Alder cycloisomerization reaction of triyne substrates⁶. We are not aware of any single-step, bimolecular reaction in which two hydrogen atoms are simultaneously transferred from a saturated alkane. Computational studies indicate a preferred geometry with eclipsed vicinal C–H bonds in the alkane donor.

Arynes^{7–9} engage in myriad trapping reactions that functionalize adjacent *sp*-hybridized carbons in the *o*-aryne ring. We recently reported a general strategy for the formation and subsequent *in situ* trapping of benzyne by means of the hexadehydro-Diels–Alder (HDDA) reaction^{6,10,11}. The simplest imaginable variant (Fig. 1b) is the reaction of 1,3-butadiyne (2) with ethyne (1, the diyneophile) to produce *o*-benzyne (3). The free energy change for this process is computed to be exothermic by approximately 50 kcal mol^{−1} (refs 6, 12). Trapping of 3 permits the synthesis of many useful benzene derivatives (4). In practice (Fig. 1c), the HDDA cycloisomerization is effected intramolecularly simply by heating a tethered triyne substrate such as 5 to produce a fused bicyclic benzyne intermediate such as 6. Trapping leads to a highly substituted benzenoid product such as 7. In addition to the preparative value of this *de novo* generation of benzyne, the HDDA reaction provides the opportunity to uncover previously unprecedented aryne trapping modes^{13,14} (for example, the insertion of the strained benzyne into the silyl ether bond as 6 proceeds to 7). This is largely because HDDA cyclizations produce reactive benzyne intermediates in the absence of added reagents, by-products or catalysts.

Bimolecular desaturation of an alkane by concerted transfer of two vicinal hydrogen atoms to a double hydrogen atom (2H) acceptor is unprecedented (Fig. 1a). We now report a 2H transfer reaction in which a HDDA-generated benzyne simultaneously accepts two vicinal hydrogen atoms from a suitable alkane 2H donor (H–C_{sp³}–C_{sp³}–H). This gives the corresponding (oxidized) alkene and (reduced) benzenoid products. For example, when we heated triyne 8 in cyclooctane to 85 °C, the only isolated product (89%) was the reduced fluorenone derivative 10-h₂ (Fig. 2a, b). Using ¹H NMR spectroscopy, we observed that a comparable amount of cyclooctene had been formed by desaturation¹⁵ (see Fig. 3b). The only well-characterized example of benzyne reduction by means of the net addition of two hydrogen atoms is a previous study¹⁶ in which a benzyne intermediate derived from a bis-diyne-bridged,

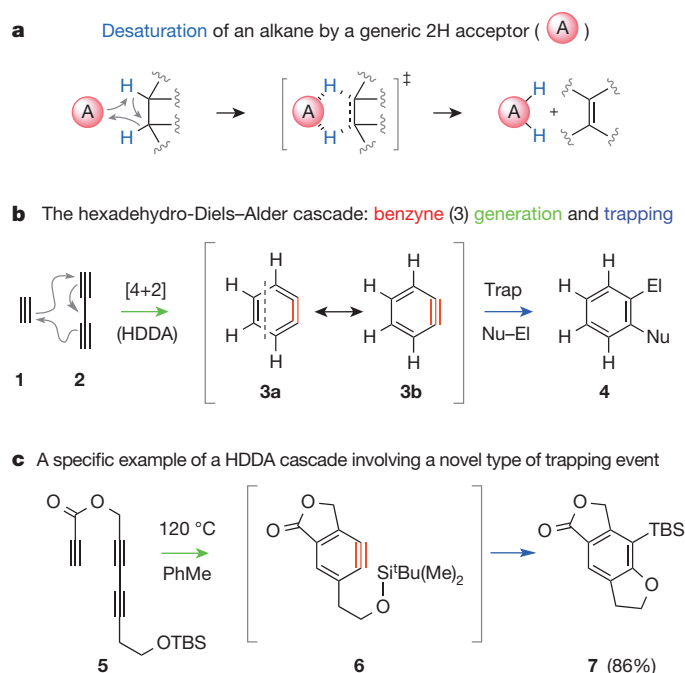


Figure 1 | Introduction and background. **a**, Species A (red sphere) is a potential double hydrogen atom (2H) acceptor, in which the accepting moiety could be either monoatomic (for example metal, metal-oxo, carbene or nitrene) or polyatomic (for example a π -bonded species) in nature. **b**, The prototypical HDDA cascade. Nu–El, nucleophile–electrophile. **c**, Intramolecular HDDA cycloisomerization followed by silyl ether trapping⁶. TBS, *tert*-butyldimethylsilyl. Here we show that a benzyne (such as 3 or 6) functions as species A by extracting two hydrogen atoms from adjacent carbon atoms of suitable 2H donor substrates.

dinuclear metal complex was reduced to the arene. They demonstrated that the solvent (tetrahydrofuran; THF) was the source of the hydrogen (and, in the case of THF-*d*₈, deuterium¹⁷) atoms that appeared in the reduced benzenoid product. When we heated substrate 8 in THF-*h*₈, 10-h₂ was the only product isolated (75%, Fig. 2a, b). Similarly, when 8 was heated in THF-*d*₈, the deuterated analogue 10-d₂ (mass spectrometry and ¹H NMR) was the only product isolated.

To probe the mechanism of this process further and, in particular, to distinguish between pathways involving sequential hydrogen atom abstractions from two solvent molecules versus a transfer of two hydrogen atoms from a single molecule, we repeated the generation and trapping of benzyne 9, this time in the presence of an equimolar mixture of THF-*h*₈ and THF-*d*₈. Intriguingly, only the diprotiobenzenoid and dideuteriobenzenoid products 10-h₂ and 10-d₂ were produced; none of the mono-H/mono-D analogue (10-hd) was detected. The observed 10-h₂:10-d₂ product ratio was 6:1, indicating a significant H/D kinetic isotope effect for the 2H transfer. In a complementary experiment, we used a 1:6 molar ratio of THF-*h*₈:THF-*d*₈, which gave a nearly 1:1 ratio of products 10-h₂:10-d₂. The lack of an observable level of monodeuterated

¹Department of Chemistry, University of Minnesota, Minneapolis, Minnesota 55455, USA.

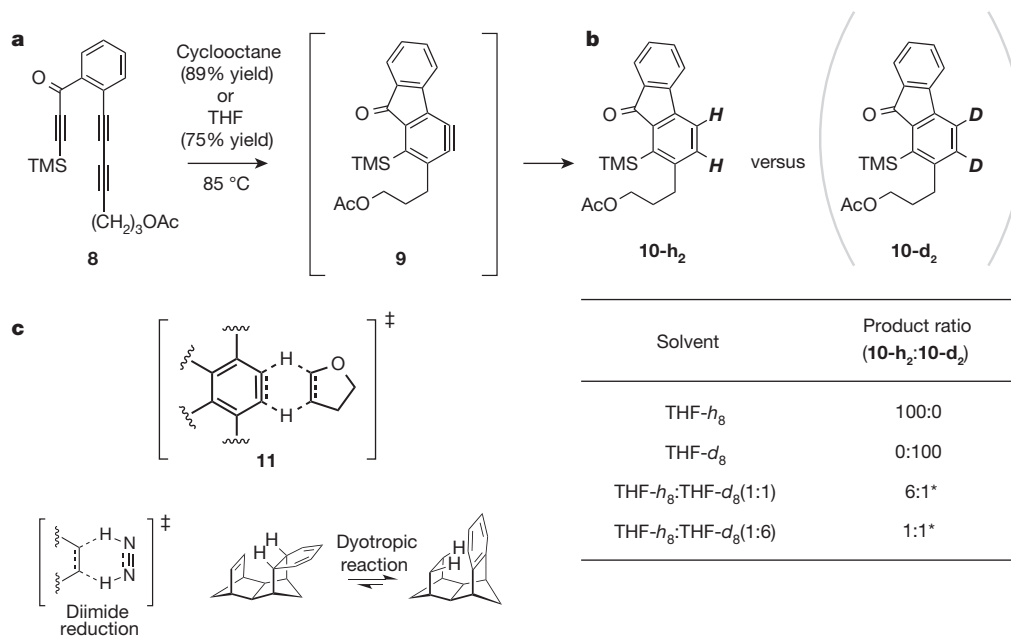


Figure 2 | Both hydrogen atoms come from the same donor molecule.

a, Dihydrogen transfer reactions between the HDDA-generated benzyne **9** (from the triene **8**) and the 2H donor solvents cyclooctane and THF give the benzenoid **10-h₂**. **b**, Isotope profiling using THF-*h*₈, THF-*d*₈ and mixtures thereof shows that both hydrogen atoms in the product originate from a single molecule of 2H donor. Asterisk, mono-deuterated product (**10-hd**) was not

product in any of these experiments is consistent with the concerted transfer to the benzyne of two hydrogen atoms from a single THF molecule as represented in the depiction of the transition structure **11** (Fig. 2c). Although such a description might seem unusual, it can be noted that the generally accepted mechanism for the reduction of alkenes by diimide (HN=NH)^{18,19} and for dyotropic reactions in which two hydrogen atoms are shuffled intramolecularly^{20,21} (Fig. 2c) involves a similar simultaneous transfer.

We next screened a series of cyclic hydrocarbons to explore their relative ability to engage an aryne in a similar hydrogen transfer reaction (Fig. 3). We were surprised to observe that cyclohexane was significantly less efficient than the other cycloalkanes in its reduction of the benzyne **13** (Fig. 3a, entries 1–5). This was initially seen from simple comparison of the chemical yields after the purification of reduced (benzenoid) product **14**. It is relevant that when the HDDA cyclization is performed in the absence of a suitably good trapping agent, we routinely observe the formation of intractable, dark-coloured mixtures of oligomeric substances. We speculate that this is because the reactive benzyne (for example **13**) engages the conjugated diyne unit in another molecule of substrate triene (for example **12**) competitively with its abstraction of two hydrogen atoms from a solvent molecule. Thus, for these processes the yield of **14** is a meaningful reflection of the 2H transfer rate from each donor solvent. Use of the acyclic hydrocarbon *n*-heptane as solvent also resulted in the production of **14**, now in 30% (isolated) yield—that is, heptane is intermediate in reactivity between cyclopentane and cyclohexane. When the starting concentration of triene **12** was decreased tenfold (that is, from 10 mM to 1 mM), the isolated yield of **14** more than doubled for the reaction in cyclohexane (from 20% to 53%) or *n*-heptane (from 30% to 73%). This is consistent with the hypothesis that the rate of the primary decomposition process is dependent on the triene concentration.

We then turned to the use of no-deuterium proton NMR (No-D NMR) spectroscopy²² to both positively identify the alkene by-product and demonstrate that it was formed in nearly equimolar amounts to the reduced benzenoid product **14**. A typical No-D spectrum, obtained with a sample prepared by heating **12** in an equivolume mixture of cyclooctane and cyclopentane ([**12**] = 0.01 M), is shown in Fig. 3b.

observed (¹H NMR analysis; estimated limit of detection was 6%).

c, Representation (**11**) of simultaneous double hydrogen transfer between an aryne and a THF molecule. Analogous six-atom arrays are involved in the transition structures of 2H transfer by diimide to an alkene acceptor^{18,19} and in the class of intramolecular reorganizations known as dyotropic reactions^{20,21}.

The ratio of the alkene resonances of cyclooctene to cyclopentene (adjusted for the molar ratio of solvents) provides the relative rate ratio (*k*_{rel}) between the two hydrocarbon donors (2.6 for this example). The relative intensity of alkene to arene resonances (from **14**) as well as the absence of resonances indicative of aromatic by-products shows the overall cleanliness of the reaction.

As stated earlier, cyclohexane (Fig. 3a, entry 5 in the inset table) is a considerably poorer 2H donor than the other cyclic hydrocarbons (entries 1–4). We speculated that this implies a preference for an eclipsed geometry for the relevant HC_{sp³}C_{sp³}H subunit within the 2H donor molecule. Accordingly, cyclohexane, dominated by the chair conformation, is least disposed towards transfer of two of its hydrogen atoms, whereas the other hydrocarbons all have low-lying conformers with HC_{sp³}C_{sp³}H dihedral angles much smaller than 60°. That is, those cyclic hydrocarbons populated to a significant extent by conformers having less highly staggered vicinal C–H bonds are the more reactive 2H donors. To test this thinking further, we examined 1,4-dioxane as a potential 2H donor. Not surprisingly, use of this chair-like compound gave none of the reduced product **14** (Fig. 3a, entry 7). In contrast, norbornane, having a boat-like cyclohexane embedded in its framework (and an associated HC_{sp³}C_{sp³}H moiety with a 0° dihedral angle) is a kinetically competent donor (entry 4), even though the product norbornene comprises a strained alkene.

Next, using density functional theory (DFT) methods (see Supplementary Information) we computed the transition structure geometry and the free energy of activation (Δ*G*[‡]) for the double hydrogen atom transfer²¹ between *o*-benzyne (**3**) and each of the seven cyclic donors shown in entries 1–7 (Fig. 3a). For all 2H donors the calculations indicate a relatively early transition structure (for example, compare the two distances shown in **15a**). This is consistent with the highly exothermic nature of the 2H transfer step (for example, we computed the free energy of reaction to be –65.6 kcal mol^{–1} for *o*-benzyne (**3**) + cyclopentane going to benzene + cyclopentene). The computed geometries for the cyclopentane and cyclohexane transition structures (**15a** and **15b**, respectively) are shown in Fig. 3c. It is not accidental that the energy difference between the chair and boat conformers of cyclohexane (approximately 6 kcal mol^{–1}) is similar to the computed difference in Δ*G*[‡] between **15a**

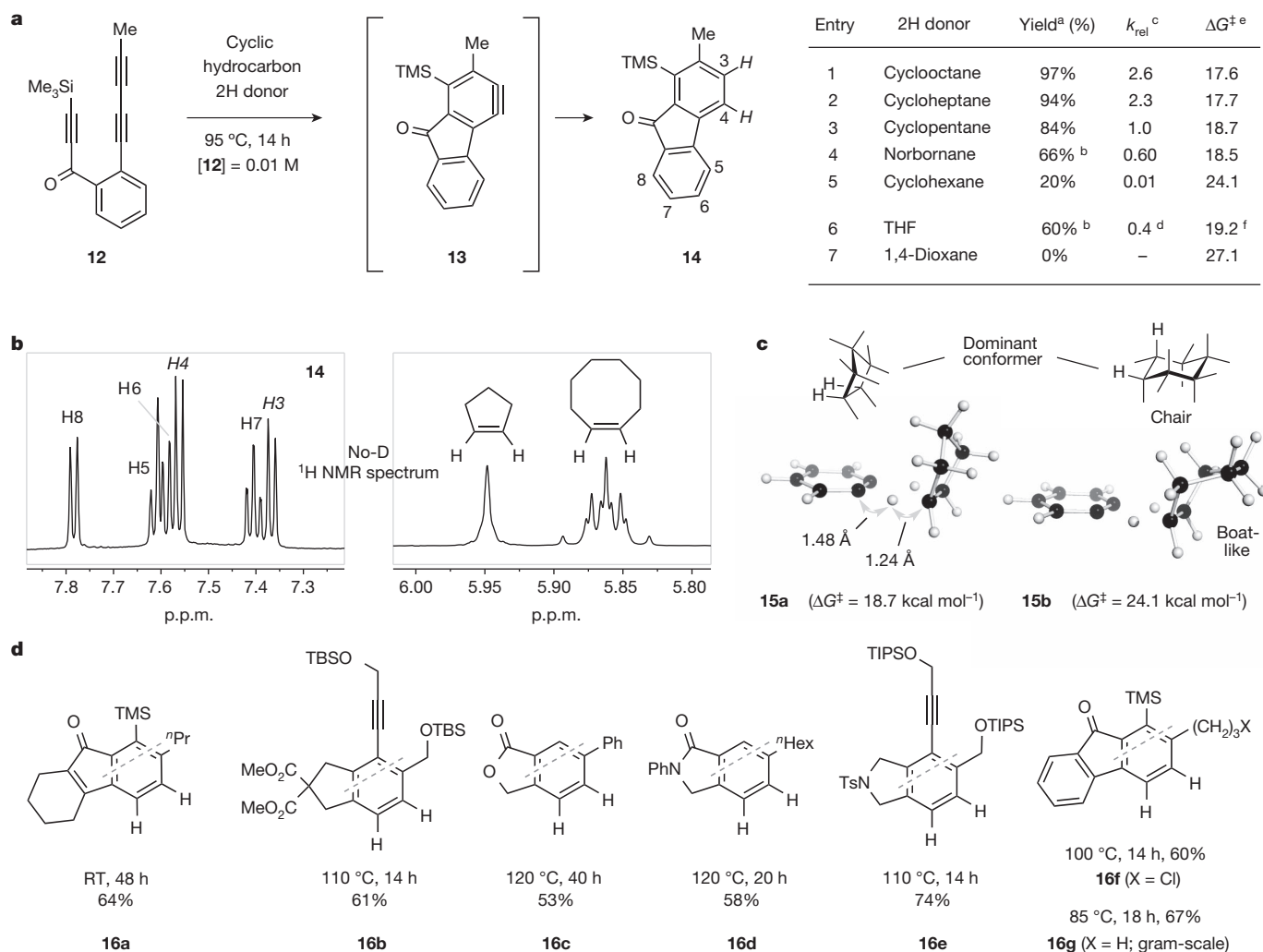


Figure 3 | Dihydrogen transfer between arynes and cyclic hydrocarbons. **a**, Relative efficiency (percentage yield and k_{rel} compared with cyclopentane) of various hydrocarbon (and cyclic ether) 2H donors for the reduction of arynes 13 to arene 14. For notes a–f in the inset table see Supplementary Information. **b**, A representative No-D ^1H NMR²² spectrum (this example is of the reaction solution arising from heating 12 (at 10 mM) in a 1.5:1 molar ratio of cyclopentane:cyclooctane at 95 °C), showing the overall efficiency of the reaction and validating the k_{rel} value (1:2.6) obtained as described in the text.

Italicized H3 and H4 denote the resonances for the newly introduced pair of hydrogen atoms in product 14. **c**, Computed transition structure geometries and ΔG^\ddagger for the transfer of two hydrogen atoms to benzyne (3) from cyclopentane (15a) and cyclohexane (15b). **d**, Reduced benzenoid products 16a–g generated by heating the triyne precursor in cyclooctane under the indicated conditions (starting substrate concentration 10 mM). RT, room temperature; Ts, *para*-toluenesulfonyl; TMS, trimethylsilyl; TIPS, triisopropylsilyl; ⁿPr, *n*-propyl; ⁿHex, *n*-hexyl.

and (the boat-like) 15b. The ΔG^\ddagger values computed for all seven donors are given in the inset table in Fig. 3a. There is a remarkably good correlation between the computed ΔG^\ddagger values and the observed k_{rel} values. These observations are most consistent with the idea of substantial dependence on dihedral angle for the process, which can only be true if the double hydrogen atom transfer event is concerted.

Products 16a–g (Fig. 3d) arose from incubating the corresponding triyne precursor (inferred from the dashed line in each structure; see Supplementary Information for details) in cyclooctane under the indicated conditions. Notable features include the following: a variety of functional groups, present in both the triyne precursor and benzenoid product, are readily tolerant of these benign reducing conditions; benzyne representing a breadth of electronic activation and/or perturbation engage in the reaction; most of the products 16 have a 1,2,3,4-tetrasubstituted motif, a substitution pattern that can be challenging to access by classical aromatic synthesis strategies; the double hydrogen atom transfer process occurs readily even at ambient temperature (compare 16a); and the reaction is not limited by scale (compare 16g).

Finally, an ancillary but important practical consideration has emerged. The most common method for generating simple benzyne derivatives,

including the parent 3, is that²³ in which 2-trimethylsilylphenyl triflate (*o*-TMSPhOTf) is exposed to a fluoride ion source (commonly CsF) in, most often, THF as the solvent. We speculated that some known trapping reactions of benzyne generated in THF are compromised in their efficiency as a result of competitive reduction by that solvent. Indeed, when we exposed *o*-TMSPhOTf to CsF in THF-*d*₈ in the absence of any other trapping agent, we observed the production of benzene (C₆H₄D₂, by ^1H NMR analysis). Similarly, benzene (and cyclopentene) was seen when CsF and *o*-TMSPhOTf were reacted in CD₃CN that contained cyclopentane (approximately 25 equivalents). We suggest that all traditional benzyne generation methods performed in the presence of a potential 2H donor (most typically, THF) are at risk to the unwanted, benzyne-depleting, 2H transfer process, especially when the benzyne trapping event is inherently slow. Indeed, we infer that this has already been encountered. Recent reports show THF to be an inferior medium (compared with 1,4-dioxane²⁴ or diethyl ether²⁵) for some benzyne trapping reactions. This is consistent with both the results reported above for the relative efficiencies of THF and 1,4-dioxane as a 2H donor (Fig. 3a, entries 6 and 7 in the inset table), and also our arguments for angle dependence during the 2H transfer.

We have described the essential mechanistic features of a double hydrogen atom transfer process. Both (vicinal) hydrogen atoms come from the same donor molecule. There is substantial dependence on dihedral angle: donors with a greater degree of eclipsing among their low-energy conformers are more reactive. This is reinforced by the nearly planar geometry of the six reacting atoms in the computed transition structures. Our observations support a pathway in which both hydrogen atoms are transferred simultaneously from the saturated alkane to the benzyne carbon atoms—a process that could be viewed as a metal-free, double C–H activation event²⁶.

METHODS SUMMARY

A typical double hydrogen atom transfer reaction comprised heating a solution of HDDA triyne precursor (substrate) in cyclooctane (approximately 0.01 M) in a closed glass reaction vessel (for example, a screw-capped vial or culture tube). After the specified time, the reaction mixture was loaded directly onto a bed of silica gel and eluted first with hexanes to remove the excess cyclooctane and then with ethyl acetate to capture the reduced benzenoid products. These were further purified by chromatography on silica gel. Relative rate data (Fig. 3a, b) were collected by ¹H NMR spectroscopy at 500 MHz using No-D NMR²² and quantitative NMR²⁷ techniques. Details are given in Supplementary Information. Details for the preparation of all new compounds, their full spectroscopic characterization data and the computational methods used are also provided in Supplementary Information.

Received 9 March; accepted 18 July 2013.

1. Linstead, R. P., Braude, E. A., Mitchell, P. W. D., Wooldridge, K. R. H. & Jackman, L. M. Transfer of hydrogen in organic systems. *Nature* **169**, 100–103 (1952).
2. Bloomfield, D. K. & Bloch, K. The formation of Δ^9 -unsaturated fatty acids. *J. Biol. Chem.* **235**, 337–345 (1960).
3. Buist, P. H. Fatty acid desaturases: selecting the dehydrogenation channel. *Nat. Prod. Rep.* **21**, 249–262 (2004).
4. Bhattacharya, A. et al. Characterization of the fungal gibberellin desaturase as a 2-oxoglutarate-dependent dioxygenase and its utilization for enhancing plant growth. *Plant Physiol.* **160**, 837–845 (2012).
5. Moran, N. A. & Jarvik, T. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* **328**, 624–627 (2010).
6. Hoye, T. R., Baire, B., Niu, D., Willoughby, P. H. & Woods, B. P. The hexadehydro-Diels–Alder reaction. *Nature* **490**, 208–212 (2012).
7. Hoffmann, R. W. *Dehydrobenzene and Cycloalkynes* (Organic Chemistry Vol. 11) (Academic, 1967).
8. Tadross, P. M. & Stoltz, B. M. A comprehensive history of arynes in natural product total synthesis. *Chem. Rev.* **112**, 3550–3577 (2012).
9. Baire, B., Niu, D., Willoughby, P. H., Woods, B. P. & Hoye, T. R. Synthesis of complex benzenoids via the intermediate generation of *o*-benzynes through the hexadehydro-Diels–Alder reaction. *Nature Protocols* **8**, 501–508 (2013).
10. Miyawaki, K., Suzuki, R., Kawano, T. & Ueda, I. Cycloaromatization of a non-conjugated polyenyne system: synthesis of 5*H*-benzo[*d*]fluoreno[3,2-*b*]pyrans via diradicals generated from 1-[2-(4-(2-alkoxymethylphenyl)butan-1,3-diynyl)]phenylpentan-2,4-diyn-1-ols and trapping evidence for the 1,2-didehydrobenzene diradical. *Tetrahedr. Lett.* **38**, 3943–3946 (1997).
11. Bradley, A. Z. & Johnson, R. P. Thermolysis of 1,3,8-nonatriyne: evidence for intramolecular [2+4] cycloaromatization to a benzyne intermediate. *J. Am. Chem. Soc.* **119**, 9917–9918 (1997).
12. Ajaz, A. et al. Concerted vs. stepwise mechanisms in dehydro-Diels–Alder reactions. *J. Org. Chem.* **76**, 9320–9328 (2011).
13. Hoffmann, R. W. & Suzuki, K. A 'hot, energized' benzyne. *Angew. Chem. Int. Ed.* **52**, 2–4 (2013).
14. Yun, S. Y., Wang, K.-P., Lee, N.-K., Mamidipalli, P. & Lee, D. Alkane C–H insertion by arynes intermediates with a silver catalyst. *J. Am. Chem. Soc.* **135**, 4668–4671 (2013).
15. Voica, A.-F., Mendoza, A., Gutekunst, W. R., Fraga, J. O. & Baran, P. S. Guided desaturation of unactivated aliphatics. *Nature Chem.* **4**, 629–635 (2012).
16. Tsui, J. A. & Sterenberg, B. T. A metal-templated 4 + 2 cycloaddition reaction of an alkyne and a diyne to form a 1,2-aryne. *Organometallics* **28**, 4906–4908 (2009).
17. de Almeida, G., Townsend, L. C. & Bertozzi, C. R. Synthesis and reactivity of dibenzoselenacycloheptynes. *Org. Lett.* **15**, 3038–3041 (2013).
18. Hünig, S., Müller, H. & Thier, W. Reduktionen mit diimid. *Tetrahedr. Lett.* **2**, 353–357 (1961).
19. Corey, E. J., Pasto, D. J. & Mock, W. L. Chemistry of diimide. II. Stereochemistry of hydrogen transfer to carbon–carbon multiple bonds. *J. Am. Chem. Soc.* **83**, 2957–2958 (1961).
20. Fernández, I., Cossío, F. P. & Sierra, M. A. Dyotropic reactions: mechanisms and synthetic applications. *Chem. Rev.* **109**, 6687–6711 (2009).
21. Fernández, I., Sierra, M. A. & Cossío, F. P. In-plane aromaticity in double-group transfer reactions. *J. Org. Chem.* **72**, 1488–1491 (2007).
22. Hoye, T. R., Eklov, B. M., Ryba, T. D., Voloshin, M. & Yao, L. J. No-D NMR (no deuterium proton NMR) spectroscopy: a simple yet powerful method for analyzing reaction and reagent solutions. *Org. Lett.* **6**, 953–956 (2004).
23. Himeshima, Y., Sonoda, T. & Kobayashi, H. Fluoride-induced 1,2-elimination of *o*-trimethylsilylphenyl triflate to benzyne under mild conditions. *Chem. Lett.* **12**, 1211–1214 (1983).
24. Ma, Z.-X., Feltenberger, J. B. & Hsung, R. P. Total syntheses of chelidonine and norchelidonine via an enamide–benzyne–[2+2] cycloaddition cascade. *Org. Lett.* **14**, 2742–2745 (2012).
25. Sumida, Y., Kato, T. & Hosoya, T. Generation of arynes via ate complexes of arylboronic esters with an *ortho*-leaving group. *Org. Lett.* **15**, 2806–2809 (2013).
26. Davies, H. M. L., Du Bois, J. & Yu, J.-Q. C–H functionalization in organic synthesis. *Chem. Soc. Rev.* **40**, 1855–1856 (2011).
27. Paul, G. F., Jak, B. & Lankin, D. A. Routine experimental protocol for qHNMR illustrated with taxol. *J. Nat. Prod.* **70**, 589–595 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank C. J. Cramer for helpful discussions about the computational studies. D.N. and P.H.W. thank the University of Minnesota Graduate School Doctoral Dissertation Fellowship and National Science Foundation Graduate Research Fellowship program, respectively. Financial support from the National Institute of General Medical Sciences (GM65597) and the National Cancer Institute (CA76497) of the US Department of Health and Human Services is acknowledged. Portions of this work were performed with hardware and software resources available through the University of Minnesota Supercomputing Institute (MSI).

Author Contributions D.N. made the initial key observations and performed most of the scope studies. P.H.W. performed most of the mechanistic studies. B.P.W. and B.B. also performed aspects of the experimental work. All authors interpreted the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.R.H. (hoye@umn.edu).

Atmospheric oxygenation three billion years ago

Sean A. Crowe^{1†*}, Lasse N. Døssing^{1,2*}, Nicolas J. Beukes³, Michael Bau⁴, Stephanus J. Kruger³, Robert Frei² & Donald E. Canfield¹

It is widely assumed that atmospheric oxygen concentrations remained persistently low (less than 10^{-5} times present levels) for about the first 2 billion years of Earth's history¹. The first long-term oxygenation of the atmosphere is thought to have taken place around 2.3 billion years ago, during the Great Oxidation Event^{2,3}. Geochemical indications of transient atmospheric oxygenation, however, date back to 2.6–2.7 billion years ago^{4–6}. Here we examine the distribution of chromium isotopes and redox-sensitive metals in the approximately 3-billion-year-old Nsuze palaeosol and in the near-contemporaneous Ijzermyn iron formation from the Pongola Supergroup, South Africa. We find extensive mobilization of redox-sensitive elements through oxidative weathering. Furthermore, using our data we compute a best minimum estimate for atmospheric oxygen concentrations at that time of 3×10^{-4} times present levels. Overall, our findings suggest that there were appreciable levels of atmospheric oxygen about 3 billion years ago, more than 600 million years before the Great Oxidation Event and some 300–400 million years earlier than previous indications for Earth surface oxygenation.

We searched for signs of early atmospheric oxygenation in rocks from the Pongola Supergroup (Fig. 1), deposited some 3 Gyr ago during the Mesoarchaeon era^{7,8}. Metamorphism in the Pongola Supergroup is largely restricted to greenschist facies^{7,8}, and Pongola rocks generally retain well-preserved geochemical signatures^{7,9} (Supplementary Information). We analysed the distribution of Cr isotopes and redox-sensitive metals both in a newly discovered palaeoweathering horizon¹⁰, the Nsuze palaeosol, formed between 2.98 and 2.96 Gyr ago, and in marine chemical sediments from the regionally expansive, shallow-water Ijzermyn iron formation^{7,9}, deposited between 2.96 and 2.92 Gyr ago (Fig. 1).

We analysed Cr isotopes in these rocks because they provide a sensitive indicator for oxidative weathering^{6,11,12} (see Supplementary Information for a detailed description of the Cr isotope proxy). This is because isotopic fractionation occurs when oxygen induces Cr redox reactions^{6,13,14}. Oxidation of trivalent Cr (Cr(III)) to soluble hexavalent Cr (Cr(VI)) can enrich the resulting Cr(VI) in the heavy ^{53}Cr isotope by comparison with the residual Cr(III) (ref. 14). In the weathering environment, this leaves soils depleted in ^{53}Cr when the ^{53}Cr -enriched Cr(VI) pool is removed by run-off¹². Partial reduction of Cr(VI) in groundwater further enriches ^{53}Cr in the mobile Cr(VI) pool, whereas light, ^{53}Cr -depleted Cr(III) is precipitated and retained¹³. The overall effect of Cr redox reactions in the weathering environment leads to heavy Cr(VI) (ref. 12) that is ultimately exported to the oceans by rivers. Chemical sediments, such as iron formations, can capture this signal of continental oxidative weathering if the mobile ^{53}Cr -enriched Cr(VI) pool is large enough to alter the Cr isotope composition of sea water⁶. Importantly, acid dissolution of Cr_2O_3 to aqueous Cr(III) does not induce measureable Cr isotope fractionation (Supplementary Information), and Cr(III) liberation by the acid weathering of rocks, which may have occurred on the continents with exceptional vigour in association with the Great Oxidation Event¹⁵, is therefore not expected to cause Cr isotope fractionation.

The oxidation of Cr(III) in soils requires the presence of manganese oxide phases^{16,17}. Manganese oxide production, in turn, requires both

molecular oxygen (O_2) and microorganisms, which catalyse the reaction between oxygen and reduced Mn (ref. 18). Overall, fractionated Cr isotopes in soils and marine sediments provide a record of oxidative weathering on the continents and the presence of O_2 in the surface environment. As an additional proxy, we also explored the distribution of uranium, which is oxidized from U(IV), forming minerals with low solubility, to soluble U(VI) in the presence of oxygen¹⁹.

We begin our exploration with the Nsuze palaeosol, which meets all five of the criteria for a true palaeoweathering profile (see additional description in the Supplementary Information), providing a new record of continental weathering in the Mesoarchaeon era¹⁰. The Nsuze palaeosol is regionally developed on basaltic andesites and marks the erosional surface that separates the basaltic andesites of the Nsuze Group from marine and fluvial sedimentary rocks of the overlying Mozaan Group¹⁰. The palaeosol is divided into two horizons on the basis of the dominant mineralogy: a thin upper sericitic horizon and a thick lower chloritic horizon (Fig. 2 and Supplementary Tables 1 and 2). The uppermost layers of the original soil have been lost, and the true thickness of the palaeosol is therefore unknown¹⁰. Extensive weathering is indicated by large depletions of Ca and Na (Supplementary Information and Supplementary Fig. 1), which extend well into the chloritic zone. High values for standard indices of alteration (CIA-K, defined as $\text{Al}_2\text{O}_3/(\text{Al}_2\text{O}_3 + \text{CaO} + \text{Na}_2\text{O}) \times 100$; ref. 20) thus reflect this massive loss of readily weathered soil components (Fig. 2). Primary igneous minerals (feldspar and amphibole) are present in the deepest samples of the lower palaeosol (Supplementary Table 2), which have elevated Ca contents and low CIA-K; therefore, this region probably approaches, but may not quite reach, the composition of the parent basaltic andesites.

Chromium isotopes are ^{53}Cr -depleted and cations are lost in a region of extensive weathering within the middle and upper reaches of the chloritic palaeosol, demonstrating Cr oxidation during weathering (Fig. 2). There is some evidence for the oxidative loss of U in the upper chloritic zone (Fig. 2), but loss based on ratios of U to some immobile elements is not as clear (Supplementary Fig. 2). Oxidative U loss may not be expected if U in the parent basaltic andesite is associated with refractory phases, such as apatites and sphene or minor zircons⁵. We also note that Fe(III) is retained in the middle and upper chlorite zones (Fig. 2). Retention of Fe(III) during primary subaerial weathering implies soil development in the presence of oxygen, whereas substantial Fe(III) loss can be expected under anoxic conditions²¹. Together, these data are best explained by weathering of the Nsuze basaltic andesites in the presence of oxygen.

We also note Fe(II) enrichment in the region of maximum weathering within the chloritic palaeosol (Fig. 2). Such enrichment could result from Fe redistribution in association with primary weathering through leaching of Fe from overlying soil horizons, as has been observed in palaeosols forming after the Great Oxidation Event²¹, when oxygen levels were substantially higher than at earlier times. There could also have been mobilization of Fe(II) during burial of the palaeosol and in the presence of anoxic groundwater. Even so, Cr and U signals derived from oxidative processes are likely to remain stable under anoxic groundwater

¹Institute of Biology, and Nordic Center for Earth Evolution, University of Southern Denmark, Odense 5230, Denmark. ²Department of Geosciences and Natural Resource Management, and Nordic Center for Earth Evolution, University of Copenhagen, Copenhagen 1350, Denmark. ³Department of Geology, University of Johannesburg, Johannesburg 2006, South Africa. ⁴Earth and Space Sciences Program, Jacobs University Bremen, Bremen 28759, Germany. [†]Present address: Department of Microbiology and Immunology, and Department of Earth, Ocean and Atmospheric Sciences, Life Sciences Centre, University of British Columbia, 2350 Health Sciences Mall, Vancouver, British Columbia V6T 1Z3, Canada.

*These authors contributed equally to this work.

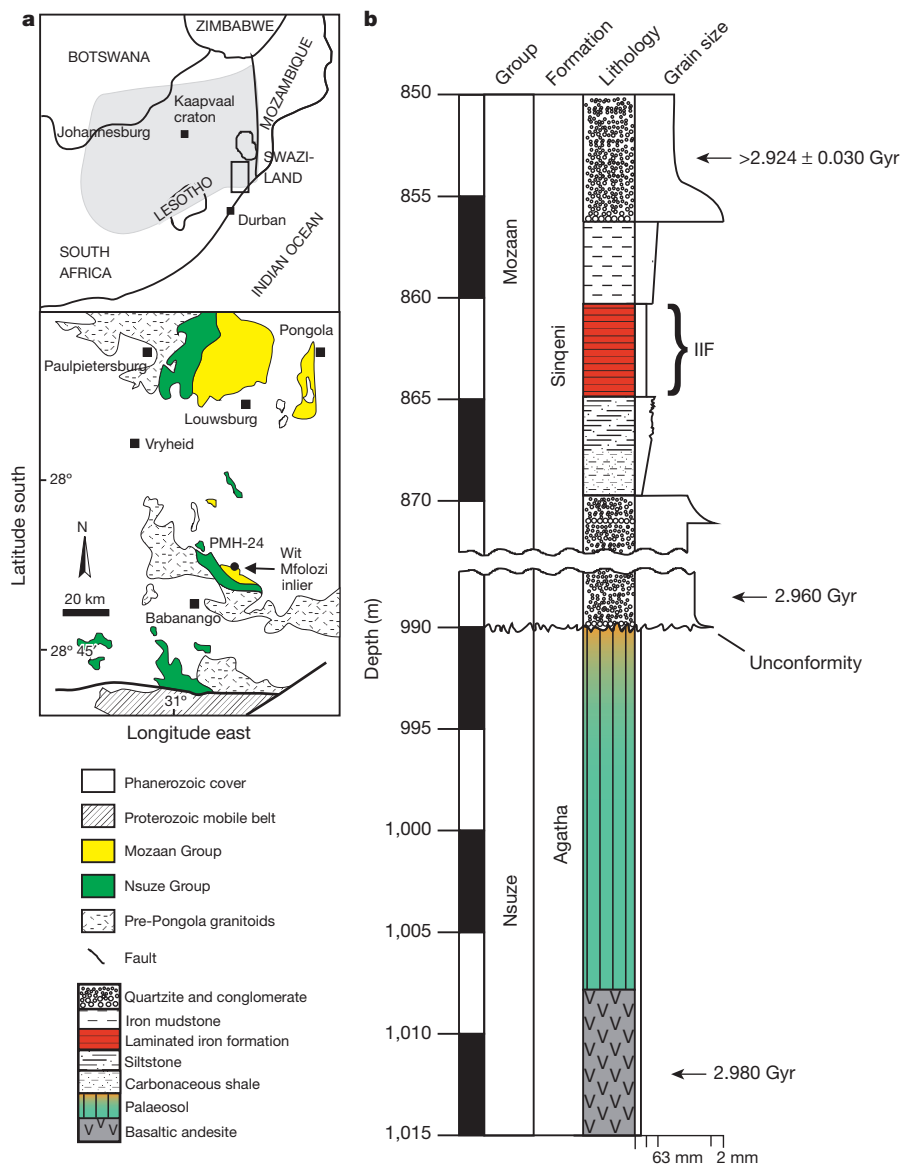


Figure 1 | Geographical location and stratigraphy. **a**, Geological map of the study area (bottom), showing the location of the Mozaan and Nsuze groups, South Africa, including the drill core PMH-24 and the outcrop of the IIF in the Wit Mfolozi inlier, and the location of the study area within South Africa (top, rectangle shows study area). **b**, Stratigraphic column of the Mozaan and Nsuze groups. Age constraints are based on previously published work for the Agatha lavas (2.980 Gyr old¹⁰), lower quartzite and conglomerate (2.960 Gyr old⁷), and upper quartzite and conglomerate (2.924 \pm 0.030 Gyr old¹⁰). On the basis of stratigraphic considerations, including the erosion of Nsuze volcanics before deposition of the lower quartzite, our best estimate for the discordance between the Nsuze palaeosol and the IIF is between 5 and 10 Myr.

conditions, unless they are overprinted by subsequent addition of Cr and U, which could occur near oxic–anoxic interfaces owing to reductive precipitation of Cr(VI) and U(VI). This seems to be the case in the upper sericitic zone of the palaeosol.

Indeed, the upper sericitic Nsuze palaeosol records fractionated Cr isotopes (though less fractionated than in the lower palaeosol) along with enrichment of Cr and U, clearly demonstrating overprinting by a later process, most likely by the reductive precipitation of Cr and U mobilized from elsewhere. Detrital processes can lead to the enrichment of elements hosted by minerals resistant to chemical weathering, but these processes are not expected to cause the fractionation of one element hosted by resistant minerals from another. Ratios of resistant elements therefore typically remain constant during detrital enrichment, and deviations from a constant ratio would indicate selective mobility induced by chemical weathering. Vertical profiles of the ratio of U to Al₂O₃ clearly show enrichment of U relative to Al₂O₃ in the sericitic palaeosol (Fig. 2e) by comparison with the parent basaltic andesite, thus indicating chemical mobility and oxidative U cycling. Redox-induced U mobility is also supported by vertical profiles (Supplementary Fig. 2) and cross plots (Supplementary Fig. 16) of U with a suite of elements commonly resistant to chemical weathering. The Cr in the sericite horizon is ⁵³Cr-enriched by comparison with the underlying chloritic horizon but is still ⁵³Cr-depleted by comparison with the

igneous inventory. Such an isotopic composition would be expected if the primary ⁵³Cr-depleted weathering signal were overprinted by ⁵³Cr-enriched Cr added through reduction of a marine, riverine or ground-water Cr(VI) pool. Importantly, though, the initial mobilization of the Cr and U, which we now see enriched in the upper sericitic palaeosol, required transport as oxidized species, and, thus, the presence of oxygen. The processes that caused the reductive precipitation of Cr and U are uncertain, but precipitation might have been related to the deposition of the overlying fluvial and marine sediments⁷. This deposition is dated to 2.96 Gyr ago⁷, and although the Cr and U chemistry of the sericite zone is complex, it still requires, albeit indirectly, atmospheric oxygen for its development.

We now turn to the near-contemporaneous Ijzermyn iron formation^{7,9} (IIF). This chemical sediment is subdivided into two units: a lower, almost exclusively chemical precipitate (CIF) and an upper, silicate facies iron formation with some siliciclastic components (SIF), as delineated by Al concentrations (Fig. 3). The iron formation is interlayered with shales⁹. Chromium in the IIF, most notably from the CIF, has $\delta^{53}\text{Cr}$ values up to $0.28 \pm 0.14\text{‰}$, with a mean for the CIF of $0.16 \pm 0.15\text{‰}$, significantly ($P_{\text{two-tailed}} = 0.0026$) heavier than the igneous inventory²² (Fig. 3). The sediments of the IIF have thus captured a mobile, ⁵³Cr-enriched Cr(VI) pool originating from oxidative continental weathering. The isotopically heavy Cr in the IIF is also accompanied by substantial

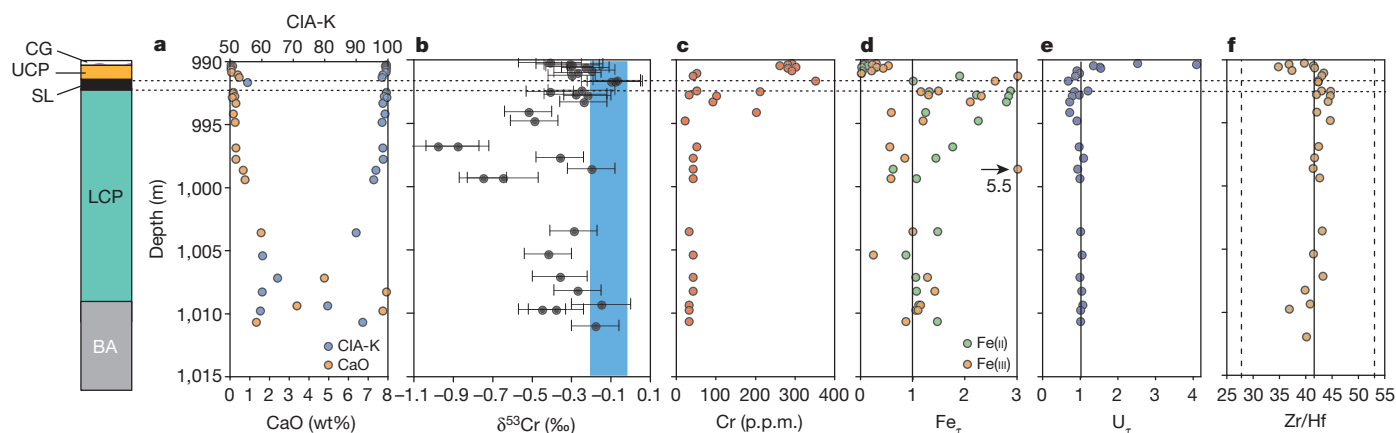


Figure 2 | Geochemical profiles and lithologies of the Nsuze palaeosol. **a**, CaO and CIA-K; **b**, $\delta^{53}\text{Cr}$; **c**, Cr concentrations; **d**, **e**, enrichment factors $\text{Fe}(\text{II})_e$ and $\text{Fe}(\text{III})_e$ (**d**) and U_e (**e**) ($\text{species}_e = [\text{species}/\text{Al}_2\text{O}_3]_{\text{sample}}/[\text{species}/\text{Al}_2\text{O}_3]_{\text{parent}}$, where $[\text{species}/\text{Al}_2\text{O}_3]_{\text{parent}}$ is calculated as the mean of the four samples from the lower chloritic palaeosol with the lowest CIA-K); **f**, Zr/Hf. The blue shaded area in **b** represents igneous-inventory $\delta^{53}\text{Cr}$ values (mean $\delta^{53}\text{Cr}_{\text{igneous}} = -0.123 \pm 0.102\text{‰}$ (2 s.d.); ref. 22); the error bars represent the in-run precision (2 s.d.) unless the in-run precision was better than the external

reproducibility (2 s.d. = 0.12‰), in which case we report the latter. The vertical dashed lines in **f** represent 30% of the mean of the Zr/Hf ratios that bracket the range exhibited by single modern weathering profiles on homogenous parent rocks²⁰. Horizontal dashed lines delineate the position of a mafic sill that crosscut the palaeosol in the Jurassic (0.228 Gyr ago) (Supplementary Information). BA, basaltic andesite; CG, conglomerate; LCP, lower chloritic palaeosol; SL, sill; USP, upper sericite palaeosol.

enrichment in U in comparison with average Archaean continental crust²³ (Fig. 3) and interlayered shales⁹ requiring a mobile U(VI) pool. The partial oxidative weathering of detrital uraninites, accumulated earlier in the Archaean eon¹⁹, provides a likely source of U(VI) for the Mesoarchaean oceans, with perhaps some input from weathered soils. Both Cr isotopes and U enrichment in the IIF therefore provide firm evidence for mobile Cr(VI) and U(VI) and signal the development of oxidative weathering in the Mesoarchaean. Palaeosol formation and deposition of the IIF may be temporally separated by up to 10 Myr, suggesting sustained oxidative weathering over this time span. Importantly, oxidative weathering in the Mesoarchaean led to a $\delta^{53}\text{Cr}$ range of 1.26‰ in the Pongola rocks, which is more than six times the range observed in all igneous rocks measured so far²². Overall, both on land and in the oceans, from source to sink, our Cr isotope and U data reveal extensive mobilization of redox-sensitive elements through oxidative weathering, implying appreciable levels of atmospheric oxygen at about 3.0 Gyr ago.

Iron formations from the 3.8-Gyr-old Isua belt in Greenland have an igneous-inventory Cr isotope composition⁶ and U abundances near the crustal inventory. These rocks provide no evidence for 3.8-Gyr-old oxidative Cr and U cycling, and a suite of iron formations ranging in age from 3.51 to 2.74 Gyr also show no Cr isotope fractionation⁶. In this record, however, there is a notable gap between 3.51 and 2.85 Gyr ago, except for a single Cr isotope measurement with a mantle-inventory value at 2.9 Gyr ago. Indeed, Cr isotope fractionation comparable to what we observed in the IIF is not seen again until 2.7–2.6 Gyr ago. This

variability in iron-formation Cr isotope composition seems to record fluctuating levels of atmospheric oxygenation during the Archaean, indicating that we captured an interval at 3.0 Gyr ago of elevated oxygen concentrations above a generally lower background level.

We can use our data to constrain levels of atmospheric oxygen 3.0 Gyr ago. Our oxygen estimates are based on the rationale that the export of Cr(VI) from the weathering environment and to the oceans requires run-off containing little Fe(II) given that Fe(II) is an effective reductant for Cr(VI), causing its precipitation and immobilization^{24,25}. Therefore, we have computed the minimum levels of atmospheric O_2 required to oxidize reactive Fe(II) during the weathering of average Archaean crust²³ (see Supplementary Information for computation details). Different assumptions about continental erosion rates and the transport rate of oxygen into soils generate different values for atmospheric O_2 . Other rock components such as Mn and redox-sensitive trace elements also act as oxygen sinks, but these are insignificant given both the dominance of Fe(II) and the order-of-magnitude uncertainty in our estimated O_2 values. Taken together, our computations suggest that oxygen concentrations were at least 6×10^{-5} times present atmospheric levels (PAL) and possibly as high as 3×10^{-3} PAL, with our best minimum estimate on the order of 3×10^{-4} PAL.

Independent atmospheric modelling of the mass-independent sulphur isotope fractionation (MIF) signal, which is small but still apparent in similarly aged Mesoarchaean rocks^{26,27}, suggests atmospheric oxygen levels below 10^{-5} PAL²⁸. These levels are similar to our lowest estimates

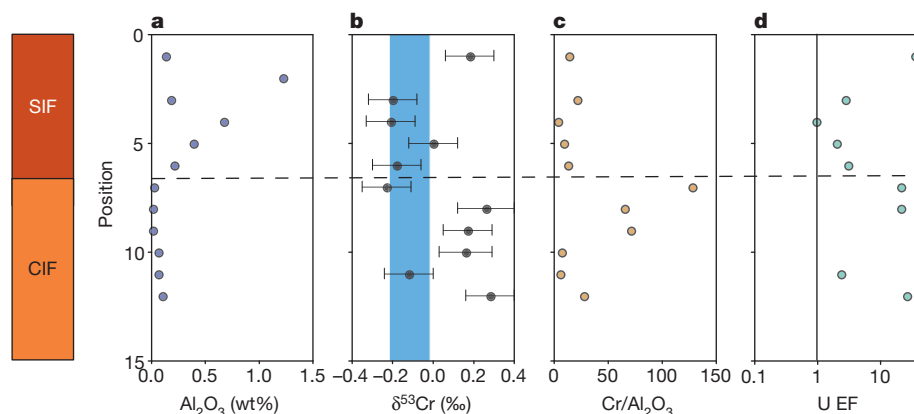


Figure 3 | Geochemical profiles of the IIF. **a**, Al_2O_3 ; **b**, $\delta^{53}\text{Cr}$; **c**, $\text{Cr}/\text{Al}_2\text{O}_3$; **d**, U enrichment factor ($U_{\text{EF}} = [\text{U}/\text{Al}_2\text{O}_3]_{\text{sample}}/[\text{U}/\text{Al}_2\text{O}_3]_{\text{crust}}$; Archaean crust $\text{U}/\text{Al}_2\text{O}_3 = 0.1412$ (ref. 23), interlayered shales $\text{U}/\text{Al}_2\text{O}_3 = 0.115$ (ref. 9)), showing a substantial enrichment of U in the IIF relative to the crust. A horizontal dashed line separates the CIF from the SIF. The blue shaded area in **b** represents igneous-inventory $\delta^{53}\text{Cr}$ values (mean $\delta^{53}\text{Cr}_{\text{igneous}} = -0.123 \pm 0.102\text{‰}$ (2 s.d.); ref. 22); the error bars represent the in-run precision (2 s.d.) unless the in-run precision was better than the external reproducibility (2 s.d. = 0.12‰), in which case we report the latter. The vertical line in **d** represents no enrichment.

for O_2 concentrations, but our higher estimate would be difficult to reconcile with these models for MIF preservation. It has been suggested that the low magnitude of the MIF signals in 2.94–2.92-Gyr-old rocks is due to dilution with mass-dependent sulphur, possibly derived from pyrite weathering²⁷. The increased levels of contemporaneous atmospheric oxygen revealed here could indeed link the low-magnitude Mesoarchean MIF signals to enhanced oxidative pyrite weathering on land and increased fluxes of sulphate to the oceans. Likewise, previous studies have also linked the preservation of detrital pyrites and uraninites to low concentrations of atmospheric oxygen in the Archaean^{1,19}. Various attempts to calibrate O_2 concentrations that allow detrital pyrite and uraninite preservation have led to the acceptance of a wide range of possible O_2 concentrations, from 10^{-2} PAL for uraninite¹⁹ to between 10^{-5} PAL⁵ and 10^{-3} PAL²⁹ for pyrite. The preservation of detrital pyrites and uraninites in the Mesoarchean is therefore well in line with the 10^{-4} PAL atmospheric O_2 proposed here.

O_2 can be produced both abiotically through atmospheric photochemical reactions and biologically through oxygenic photosynthesis. Atmospheric modelling³⁰ has placed upper limits of 2.5×10^{-9} PAL on photochemical oxygen production via hydrogen peroxide intermediates, although, under very special circumstances, up to 1.5×10^{-5} PAL can be produced locally. Our observations require oxidative weathering over broad swaths of the continents, and so even our lower estimate of 6×10^{-5} PAL oxygen greatly exceeds those possible by photochemical means alone. We suggest oxygen production through oxygenic photosynthesis as a possible explanation for the geochemical signals we observe in 3.0-Gyr-old rocks. This would imply that cyanobacteria may have evolved by this time, indicating a much deeper history, by 300–400 Myr, for oxygenic photosynthesis than previously realized¹.

METHODS SUMMARY

Samples of the Nsuzi palaeosol were recovered from drill core PMH-24, whereas unweathered samples of the IIF were collected from outcrops within the Wit Mfolozi inlier. All samples were crushed to powder using an agate mortar. Before Cr purification for isotope measurements, samples were spiked with a Cr(III) ^{50}Cr – ^{54}Cr isotope double spike⁸ and dissolved in hot HF – HNO_3 – HCl acid mixtures. Chromium was purified using the standard ion exchange technique and the isotope ratios measured by thermal ionization mass spectrometry, as previously described⁸. Chromium isotope ratios are reported as

$$(\delta^{53}\text{Cr} = (^{53}\text{Cr}/^{52}\text{Cr})_{\text{sample}} / (^{53}\text{Cr}/^{52}\text{Cr})_{\text{SRM 979}} - 1) \times 1000$$

where SRM 979 denotes Standard Reference Material 979. Trace elements were determined on a Perkin Elmer Elan 6100 DRC quadrupole inductively coupled plasma mass spectrometer using the international BHVO-2 standard for calibration. Whole-rock analyses were carried out at Activation Laboratories Ltd following the analytical procedure termed 'Code 4LITHO Major Elements Fusion ICP(WRA)/Trace Elements Fusion ICP/MS'.

Full Methods and any associated references are available in the online version of the paper.

Received 12 October 2012; accepted 28 June 2013.

- Canfield, D. E. The early history of atmospheric oxygen: homage to Robert A. Garrels. *Annu. Rev. Earth Planet. Sci.* **33**, 1–36 (2005).
- Holland, H. D. Volcanic gases, black smokers, and the Great Oxidation Event. *Geochim. Cosmochim. Acta* **66**, 3811–3826 (2002).
- Guo, Q. J. *et al.* Reconstructing Earth's surface oxidation across the Archaean-Proterozoic transition. *Geology* **37**, 399–402 (2009).
- Wille, M. *et al.* Evidence for a gradual rise of oxygen between 2.6 and 2.5 Ga from Mo isotopes and Re-PGE signatures in shales. *Geochim. Cosmochim. Acta* **71**, 2417–2435 (2007).
- Anbar, A. D. *et al.* A whiff of oxygen before the Great Oxidation Event? *Science* **317**, 1903–1906 (2007).
- Frei, R., Gaucher, C., Poulton, S. W. & Canfield, D. E. Fluctuations in Precambrian atmospheric oxygenation recorded by chromium isotopes. *Nature* **461**, 250–253 (2009).
- Beukes, N. J. & Cairncross, B. A lithostratigraphic sedimentological reference profile for the Late Archaean Mozaan Group, Pongola Sequence: application to

sequence stratigraphy and correlation with the Witwatersrand Supergroup. *S. Afr. J. Geol.* **94**, 44–69 (1991).

- Mukasa, S. B., Wilson, A. H. & Young, K. R. Geochronological constraints on the magmatic and tectonic development of the Pongola Supergroup (Central Region), South Africa. *Precamb. Res.* **224**, 268–286 (2013).
- Alexander, B. W., Bau, M., Andersson, P. & Dulski, P. Continentally-derived solutes in shallow Archean seawater: Rare earth element and Nd isotope evidence in iron formation from the 2.9 Ga Pongola Supergroup, South Africa. *Geochim. Cosmochim. Acta* **72**, 378–394 (2008).
- Nhleko, N. *The Pongola Supergroup in Swaziland*. PhD thesis, Rand Afrikaans Univ. (2003).
- Frei, R. & Polat, A. Chromium isotope fractionation during oxidative weathering—Implications from the study of a Paleoproterozoic (ca. 1.9 Ga) paleosol, Schreiber Beach, Ontario, Canada. *Precamb. Res.* **224**, 434–453 (2013).
- Crowe, S. A. *et al.* Oxidative weathering fractionates chromium isotopes. *Mineral. Mag.* **75**, 706 (2011).
- Ellis, A. S., Johnson, T. M. & Bullen, T. D. Chromium isotopes and the fate of hexavalent chromium in the environment. *Science* **295**, 2060–2062 (2002).
- Zink, S., Schoenberg, R. & Staubwasser, M. Isotopic fractionation and reaction kinetics between Cr(III) and Cr(VI) in aqueous media. *Geochim. Cosmochim. Acta* **74**, 5729–5745 (2010).
- Konhauser, K. O. *et al.* Aerobic bacterial pyrite oxidation and acid rock drainage during the Great Oxidation Event. *Nature* **478**, 369–373 (2011).
- Eary, L. E. & Rai, D. Kinetics of chromium (III) oxidation to chromium (VI) by reaction with manganese dioxide. *Environ. Sci. Technol.* **21**, 1187–1193 (1987).
- Oze, C., Bird, D. K. & Fendorf, S. Genesis of hexavalent chromium from natural sources in soil and groundwater. *Proc. Natl Acad. Sci. USA* **104**, 6544–6549 (2007).
- Tipping, E. Temperature dependence of Mn(II) oxidation in lakewaters: a test of biological involvement. *Geochim. Cosmochim. Acta* **48**, 1353–1356 (1984).
- Grandstaff, D. E. Origin of uraniferous conglomerates at Elliot Lake, Canada, and Witwatersrand, South Africa: implications for oxygen in the Precambrian atmosphere. *Precamb. Res.* **13**, 1–26 (1980).
- Maynard, J. B. Chemistry of modern soils as a guide to interpreting Precambrian paleosols. *J. Geol.* **100**, 279–289 (1992).
- Ohmoto, H. Evidence in pre-2.2 Ga paleosols for the early evolution of atmospheric oxygen and terrestrial biota. *Geology* **24**, 1135–1138 (1996).
- Schoenberg, R., Zink, S., Staubwasser, M. & von Blanckenburg, F. The stable Cr isotope inventory of solid Earth reservoirs determined by double spike MC-ICP-MS. *Chem. Geol.* **249**, 294–306 (2008).
- Condie, K. C. Chemical composition and evolution of the upper continental crust: contrasting results from surface samples and shales. *Chem. Geol.* **104**, 1–37 (1993).
- Eary, L. E. & Rai, D. Kinetics of chromate reduction by ferrous ions derived from hematite and biotite at 25 degrees C. *Am. J. Sci.* **289**, 180–213 (1989).
- Sass, B. M. & Rai, D. Solubility of amorphous chromium(III)-iron(III) hydroxide solid-solutions. *Inorg. Chem.* **26**, 2228–2232 (1987).
- Farquhar, J. *et al.* Isotopic evidence for Mesoarchean anoxia and changing atmospheric sulphur chemistry. *Nature* **449**, 706–709 (2007).
- Guy, B. M. *et al.* A multiple sulfur and organic carbon isotope record from non-conglomeratic sedimentary rocks of the Mesoarchean Witwatersrand Supergroup, South Africa. *Precamb. Res.* **216–219**, 208–231 (2012).
- Pavlov, A. A. & Kasting, J. F. Mass-independent fractionation of sulfur isotopes in Archaean sediments: strong evidence for an anoxic Archaean atmosphere. *Astrobiology* **2**, 27–41 (2002).
- Canfield, D. E., Habicht, K. S. & Thamdrup, B. The Archean sulfur cycle and the early history of atmospheric oxygen. *Science* **288**, 658–661 (2000).
- Haqq-Misra, J., Kasting, J. F. & Lee, S. Availability of O_2 and H_2O_2 on pre-photosynthetic Earth. *Astrobiology* **11**, 293–302 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements N. Planavsky, R. Schoenberg, S. Poulton, A. Basu, C. Jones, H. Tsikos, A. Mucci, A. O'Neill and T. Dahl are thanked for suggestions. T. Larsen, C. N. Jensen, T. Leeper and P. Søholt are acknowledged for technical support. Funding to S.A.C. was provided by an Agouron Institute Geobiology Fellowship and an NSERC PDF. Additional funding was from the Danish National Research Foundation (grant no. DNR53), the Danish Agency for Science, Technology, and Innovation, the European Research Council and the National Research Foundation in Pretoria. The palaeosol drill core was made available by Ian Frith of AngloGold Ashanti Exploration (SA), from their core store in Carltonville, South Africa.

Author Contributions S.A.C., N.J.B. and L.N.D. had the idea for the study; samples were provided by N.J.B. and M.B.; Cr isotope measurements were made by L.N.D.; other geochemical analyses were performed by L.N.D., S.A.C., N.J.B. and S.J.K.; S.A.C., L.N.D. and D.E.C. produced the manuscript with significant contributions from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.A.C. (sean.crowe@ubc.ca).

METHODS

Samples were collected from a drill core (PMH-24) in the Denny Dalton area¹⁰, South Africa, and from outcrops of the IIF, Singeni Formation of the Mozaan Group⁹, which outcrops within the Wit Mfolozi inlier (Fig. 1a and Supplementary Fig. 3). All samples for elemental measurements were crushed to powder using an agate mortar. Whole-rock analyses were carried out at Activation Laboratories Ltd following the analytical procedure termed 'Code 4LITHO Major Elements Fusion ICP(WRA)/Trace Elements Fusion ICP/MS'. A brief description of the procedure is summarized below from information available on the Actlabs homepage (<http://www.actlabs.com>). Crushed samples were mixed with a flux of lithium metaborate and lithium tetraborate and fused in an induction furnace. The molten melt was immediately poured into a solution of 5% nitric acid containing an internal standard, and mixed continuously until completely dissolved (~30 min).

For 'Code 4B', the samples were analysed for major oxides and selected trace elements using a combination of instruments simultaneously or sequentially (Thermo Jarrell-Ash ENVIRO II ICP or a Varian Vista 735 ICP). Calibration was performed using seven prepared USGS- and CANMET-certified reference materials. One of the seven standards was used during the analysis for every group of ten samples. For 'Code 4B2', the fused samples were diluted and analysed by inductively coupled plasma mass spectrometry (Perkin Elmer Sciex ELAN 6000). Three blanks and five controls (three before the sample group and two after) were analysed per group of samples. Duplicates were fused and analysed every 15 samples. The instrument was recalibrated every 40 samples. We note that samples were not decarbonated before analyses, and so CaO values include carbonate Ca. Carbonate minerals were, however, not detected by X-ray diffraction analyses and are therefore probably a minor component of the palaeosol.

Additionally, the Fe(II) content was determined as FeO (Actlabs code 4F-FeO-titration) by titration using a cold acid digestion with ammonium metavanadate and HF. Ferrous ammonium sulphate was added after digestion and potassium dichromate was used as the titrating agent. The end point was determined visually on the basis of colour.

Standard X-ray powder diffraction analyses were conducted on palaeosol samples to identify dominant mineral phases. Selected minerals were analysed for their microchemical composition by means of a Cameca SX100 electron microprobe. Reference material 'Multi RM block PHI' was also analysed.

Additional trace element analyses were conducted using solution inductively coupled plasma mass spectrometry (Perkin Elmer Elan 6100 DRC quadrupole ICP-MS). The samples were dissolved by standard procedures using HCl–HNO₃–HF. The international reference material BHVO-2 was analysed for calibration. In general, these data agreed well with those obtained from Actlabs.

Before purification, a ⁵⁰Cr–⁵⁴Cr double spike was added to the samples to correct for potential Cr isotope fractionation during separation as well as instrumental mass bias during analysis. A two-step Cr purification procedure was used⁶. In the first separation step, digests were passed through 10-ml Poly-Prep Chromatography Columns (Bio Rad Laboratories, no. 731-1550EDU) packed with 2 ml DOWEX AG-1x8 anion resin (100–200 mesh chloride form, Bio Rad Laboratories) that

was preconditioned with 6 mol l^{−1} HCl. In 6 mol l^{−1} HCl, Cr(III) has low affinity for the resin and passes through the column. Chromium(III) was collected with 6 ml of 6 mol l^{−1} HCl in Savillex PFA vials and evaporated to dryness on a hot plate. The resin was cleaned with 20 ml of 5 mol l^{−1} HNO₃, 30 ml of 6 mol l^{−1} HCl and 40 ml of ultrapure water, and reused in the second chromatographic step for the same sample to minimize contamination. To oxidize all Cr(III) to Cr(VI), the Cr(III) was redissolved in 10 ml of 0.2 mol l^{−1} HCl, 0.5 ml of 0.5 μmol l^{−1} (NH₄)₂S₂O₈ was added and the sample was heated to 130 °C for 1 h.

A second chromatographic step was used to separate Cr(VI) from the remaining matrix, using an anion resin that was preconditioned with 0.2 mol l^{−1} HCl. This second step is based on the exchange of chloride ions on the anion resin with Cr(VI) oxyanions as previously described⁶. After the elution of matrix elements with 10 ml 2.0 mol l^{−1} HCl, Cr was released from the anion resin and eluted by reduction to Cr(III) using 2 mol l^{−1} HNO₃ and 5% hydrogen peroxide. Finally, the Cr-bearing solutions were evaporated to dryness on a hot plate and the purified Cr mounted on Re filaments with a mixture of 3 μl silica gel, 0.5 μl 0.5 mol l^{−1} of H₃BO₃ and 0.5 μl 0.5 mol l^{−1} of H₃PO₄. Notably, all samples were subjected to both chromatographic steps except soil extracts and water samples, which were subjected only to the second step, because these samples contained low concentrations of matrix elements. Procedural yield tests were performed by passing known quantities of the certified isotope standard reference material NIST SRM 979 over chromatographic columns with double-spike addition after Cr chromatography. Chromium recovery varied between 80% and 90%. Procedural blanks contained 5–10 ng Cr, whereas a typical sample yielded >500 ng purified Cr. All Cr isotope ratio measurements were conducted using an IsotopX/GV IsoProbe T thermal ionization mass spectrometer equipped with eight Faraday collectors to monitor ⁵⁰Cr⁺, ⁵²Cr⁺, ⁵³Cr⁺, ⁵⁴Cr⁺, ⁴⁹Ti⁺, ⁵¹V⁺, ⁵⁶Fe⁺ and ⁵⁵Mn⁺. One measurement consisted of 120 cycles (grouped into 24 blocks of 5 cycles each with 10-s signal integration periods) performed in static mode. Before every block, a baseline integration of 20 s at 0.5 AMU was collected on either side of the peaks. Every load was analysed 2–7 times. The final isotope composition of a sample was determined as the average of the repeated analyses. Processed double-spiked NIST SRM 979 standard yielded a 2 s.d. external reproducibility in δ⁵³Cr value of ±0.12‰ (52 AMU signal intensity of 0.4 V).

Different separates of the crosscutting sill were obtained from crushed whole rock powders using standard magnetic separation and heavy liquid techniques. Before purification, the rock powders were spiked with a ¹⁴⁹Sm–¹⁵⁰Nd mixed spike and the bulk rare-earth elements were separated over 15-ml glass-stem columns charged with AG 50W cation resin (100–200 Mesh). Rare-earth elements were further separated over HDEHP-coated bio beads (Bio Rad Laboratories) loaded in 6 ml glass-stem columns. Samarium and Nd isotope ratios were analysed using a VG Sector 54 IT mass spectrometer. During analysis, both static and multidynamic routines were used for the collection of the isotopic ratios. Neodymium isotope ratios were normalized to ¹⁴⁶Nd/¹⁴⁴Nd = 0.7219. The mean value of ¹⁴³Nd/¹⁴⁴Nd for the JNdi-1 standard during the period of measurement was 0.512109 ± 0.000009 (2 s.d., N = 5).

Integrating abundance and functional traits reveals new global hotspots of fish diversity

Rick D. Stuart-Smith¹, Amanda E. Bates¹, Jonathan S. Lefcheck², J. Emmett Duffy², Susan C. Baker³, Russell J. Thomson¹, Jemina F. Stuart-Smith¹, Nicole A. Hill¹, Stuart J. Kininmonth⁴, Laura Airolidi^{5,6}, Mikel A. Becerro⁷, Stuart J. Campbell⁸, Terence P. Dawson⁹, Sergio A. Navarrete¹⁰, German A. Soler¹, Elisabeth M. A. Strain⁵, Trevor J. Willis¹¹ & Graham J. Edgar¹

Species richness has dominated our view of global biodiversity patterns for centuries^{1,2}. The dominance of this paradigm is reflected in the focus by ecologists and conservation managers on richness and associated occurrence-based measures for understanding drivers of broad-scale diversity patterns and as a biological basis for management^{3,4}. However, this is changing rapidly, as it is now recognized that not only the number of species but the species present, their phenotypes and the number of individuals of each species are critical in determining the nature and strength of the relationships between species diversity and a range of ecological functions (such as biomass production and nutrient cycling)⁵. Integrating these measures should provide a more relevant representation of global biodiversity patterns in terms of ecological functions than that provided by simple species counts. Here we provide comparisons of a traditional global biodiversity distribution measure based on richness with metrics that incorporate species abundances and functional traits. We use data from standardized quantitative surveys of 2,473 marine reef fish species at 1,844 sites, spanning 133 degrees of latitude from all ocean basins, to identify new diversity hotspots in some temperate regions and the tropical eastern Pacific Ocean. These relate to high diversity of functional traits amongst individuals in the community (calculated using Rao's Q^6), and differ from previously reported patterns in functional diversity and richness for terrestrial animals, which emphasize species-rich tropical regions only^{7,8}. There is a global trend for greater evenness in the number of individuals of each species, across the reef fish species observed at sites ('community evenness'), at higher latitudes. This contributes to the distribution of functional diversity hotspots and contrasts with well-known latitudinal gradients in richness^{2,4}. Our findings suggest that the contribution of species diversity to a range of ecosystem functions varies over large scales, and imply that in tropical regions, which have higher numbers of species, each species contributes proportionally less to community-level ecological processes on average than species in temperate regions. Metrics of ecological function usefully complement metrics of species diversity in conservation management, including when identifying planning priorities and when tracking changes to biodiversity values.

Species do not contribute equally to varied ecosystem processes and delivery of services⁹, and thus the sustainability and resilience of these processes depend on aspects of diversity beyond the number of species present in a community. This rationale has led to recent exploration of functional diversity indices to describe variation in traits that mediate species' contributions to ecosystem processes¹⁰. Community-wide contributions to a range of processes will depend not only on which species and functional traits are present but also on their relative abundances;

yet only a few studies of functional diversity have accounted for the uneven distribution of individuals of different species within a community¹¹. So far, these studies have not used consistent quantitative sampling to account for the functional consequences of widely differing species abundances at a global scale.

Our global analysis focuses on reef fishes, using data collected from 4,357 underwater visual transects. Fishes are ideal for a global diversity analysis, as they represent the most diverse vertebrate group, encompass a range of important and diverse ecological roles, and marine fishes are relatively well known taxonomically and functionally. We build on prior analyses of coral reef fish diversity^{12,13} by substantially expanding the reef fish database in the world's temperate and sub-polar regions, providing the first global maps of abundance-weighted functional diversity for any taxon in the oceans, and assessing congruence between richness and abundance-weighted functional diversity measures.

Our functional diversity index summarizes community diversity with respect to eight traits that contribute to the functional position of species in the ecosystem, encompassing body size (maximum length), feeding ecology (trophic group, trophic breadth), behaviour (water column position, diel activity pattern, gregariousness), and habitat use (preferred substratum, habitat complexity) (Extended Data Table 1). We calculate functional diversity using Rao's quadratic entropy⁶, Q , a measure of community-level dispersion of species in functional trait space weighted by their relative abundances. Rao's Q is not mathematically constrained to be positively correlated with species richness⁶, and thus allows unbiased tests of the relationships between species diversity and functional diversity. We use functional group richness to represent the richness of species groups sharing similar functional traits, for the purpose of describing a traditional representation of functional diversity, and for comparison with other diversity metrics.

Previous global and broad-scale studies of marine fishes based on compilations of species distribution records have shown that species richness declines with increasing latitude and distance from hotspots in southeast Asia and the Caribbean^{12–14}. Here we corroborate these patterns for reef fishes using the number of species observed in a standardized area at the scale of individual reefs ('species density'; Fig. 1a). Global mapping was based on predictions using broad-scale environmental variables, from random forest models that were generated from the site-level diversity measures and corresponding environmental data, with the effect of inherent spatial autocorrelation evaluated (see Methods). As expected from studies of other taxa⁷, the richness of functional groups (Fig. 1b) closely resembles the pattern in species density, with more trait combinations present in speciose tropical regions.

Functional diversity of reef fish communities, incorporating relative abundances of species, shows markedly different global patterns to

¹Institute for Marine and Antarctic Studies, University of Tasmania, Private Bag 49, Hobart, Tasmania 7001, Australia. ²Department of Biological Sciences, Virginia Institute of Marine Science, P.O. Box 1346, Gloucester Point, Virginia 23062-1346, USA. ³School of Plant Science, University of Tasmania, Private Bag 55, Hobart, Tasmania 7001, Australia. ⁴Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, Stockholm SE-114 19, Sweden. ⁵Dipartimento di Scienze Biologiche, Geologiche ed Ambientali, Università di Bologna, Via San Alberto 163, Ravenna 48123, Italy. ⁶Hopkins Marine Station, Stanford University, Pacific Grove, California 93950, USA. ⁷Natural Products and Agrobiological Institute (IPNA-CSIC), Avda Astrofísico Francisco Sánchez 3, La Laguna 38206, Tenerife, Spain. ⁸Wildlife Conservation Society, Indonesia Marine Program, Jalan Atletik no. 8, Bogor, Jawa Barat 16151, Indonesia. ⁹School of the Environment, University of Dundee, Perth Road, Dundee DD1 4HN, Scotland, UK. ¹⁰Estación Costera de Investigaciones Marinas & Center for Marine Conservation, Pontificia Universidad Católica de Chile, Casilla 114-D, Santiago, Chile. ¹¹School of Biological Sciences, Institute of Marine Sciences, University of Portsmouth, Ferry Road, Eastney, Portsmouth PO4 9LY, UK.

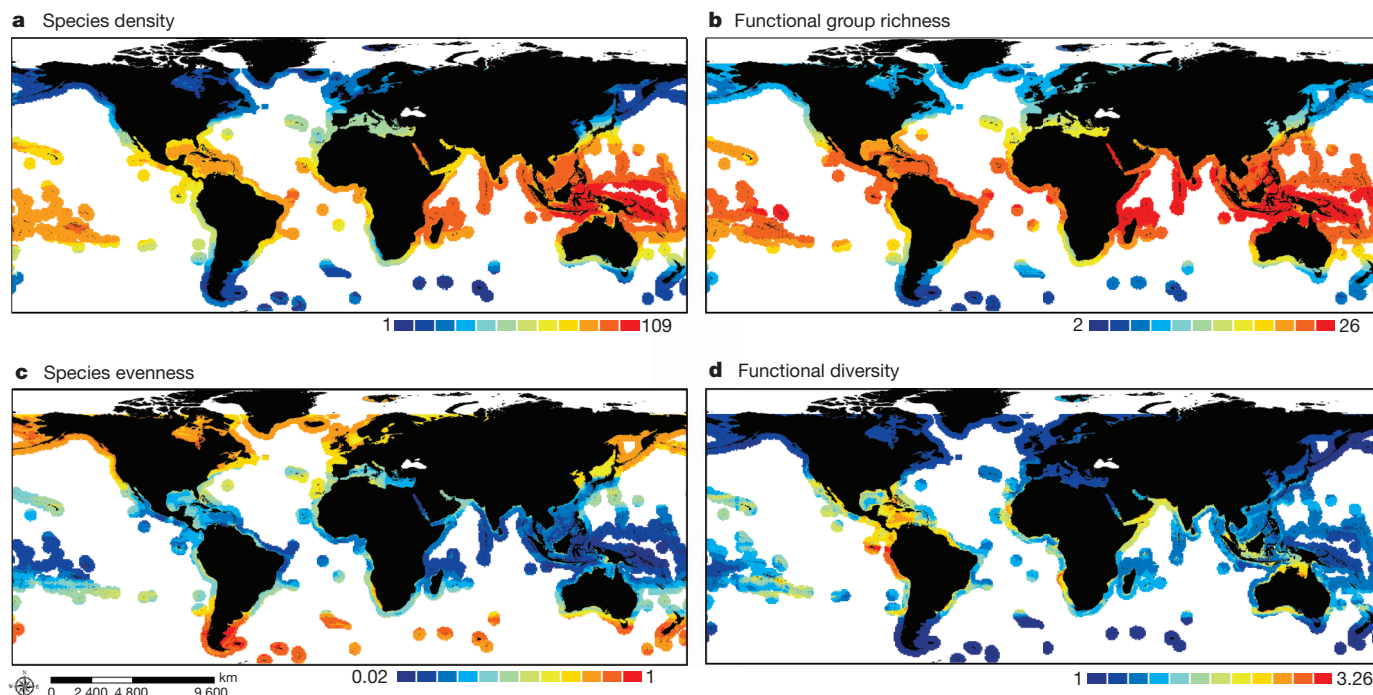


Figure 1 | Global fish diversity patterns predicted from quantitative divers censuses at 1,844 sites. a, Species density (a relative measure of species richness) matches the previously established gradient (decreasing with higher latitudes). **b**, Functional group richness mirrors species density. **c**, Evenness (the spread of the number of individuals of each species, across species)

those of species density and functional group richness (Fig. 1d), and is influenced by a previously undocumented global trend of increasing community evenness with latitude (Fig. 1c). Although current understanding of global patterns in functional diversity is based largely on richness-based measures⁷, and thus highlights only tropical regions as hotspots, we found only a weak relationship between functional group richness and abundance-weighted functional diversity ($r = 0.21$; Extended Data Fig. 1; this was weaker when functional richness was expressed as the volume of trait space occupied). It is well-documented that the ecological effects of a species are generally proportional to its abundance or biomass¹⁵. Thus, by incorporating observed variation in abundances across species within sites, abundance-weighted functional diversity more accurately reflects community functional structure than diversity metrics based on a simple count or inventory of species present within a broader geographic region. Our abundance-weighted metric provides an improved measure of functionally relevant biodiversity and an important new perspective to our understanding of global biodiversity. In addition to some tropical regions, new hotspots of fish diversity are predicted for nutrient-rich regions, such as in cool boundary currents, where plankton and macroalgal trophic pathways are prominent. Conversely, oligotrophic regions, including some that are heavily impacted such as the Mediterranean, generally have low functional diversity (Fig. 1d), regardless of whether the total number of functional groups is high.

The reasons for high functional diversity are not consistent across hotspots. Sites that were surveyed in temperate regions were generally characterized by higher functional diversity relative to species diversity (or richness) than sites in the tropics (Fig. 2 and Extended Data Table 2) and greater evenness in abundances across species (high community evenness; Fig. 1c). Thus, temperate functional diversity hotspots result from high average variation between individual fish in terms of their functional traits and likely functional roles ('high individual complementarity'). Tropical hotspots generally occur in regions where moderate to high functional group richness coincides with regions of moderate

increases with latitude. **d**, Abundance-weighted functional diversity is highest in the tropical eastern Pacific Ocean and dispersed hotspots at a range of latitudes. Colour classifications differ between maps owing to different ranges and distributions of diversity values. Minimum and maximum observed values are provided in the key for each plot as effective numbers per 500 m² (see Methods).

community evenness, and thus can result from high average differentiation (complementarity) across either species or individuals (or both). The Galapagos Archipelago, which effectively comprises an isolated nutrient-rich, temperate-like region straddling the equator¹⁶, is an exception to this pattern. This global hotspot is the result of uniquely high inter-specific complementarity, where fish communities are the most functionally diverse, worldwide, for any given level of species diversity (Fig. 2), despite relatively low evenness in abundance.

Our study also provides an important perspective on the relationship between diversity and ecosystem functioning. This has typically been explored in terms of how the number of species influences various ecosystem properties in experimental systems at small scales (see ref. 17), but has not often taken advantage of large-scale field observations^{18,19}. Specifically, by focusing on multiple traits and on natural abundances, rather than simply presence or absence of species, we reveal a picture of reef fish functional diversity that varies over geographic gradients in a pattern distinct from global trends in species or functional richness (Fig. 1). The main reason for this discrepancy is that in locations in which species have more even abundance distributions, such as within temperate functional diversity hotspots, a greater proportion of species reach moderate abundances at which their unique trait combinations can significantly influence ecological processes. In contrast, functionally unique tropical species tend to be rare, with a relatively weak average influence on ecosystem processes. However, this does not mean that rare species may not sometimes still have important ecological roles²⁰, and species with unique traits are clearly important from a conservation perspective²¹.

To develop more effective biodiversity management to maximize the reliability of ecosystem processes in a changing environment requires cost-effective, broad-scale assessment and monitoring of many dimensions of biodiversity, including consideration of functional traits²². For example, metrics of ecological function can provide an important additional input for marine protected area (MPA) planning and management. Relatively few MPAs are located at temperate latitudes, including

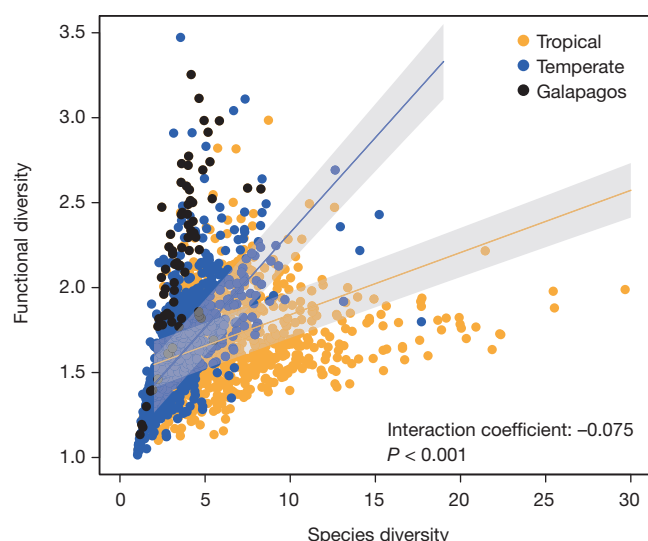


Figure 2 | The species diversity–functional diversity relationship for reef fishes differs between temperate and tropical sites. Although overlap between some temperate and tropical sites was observed, on average temperate sites have higher functional diversity relative to species diversity compared to the tropics ($P < 0.001$, $n = 1,817$ sites; see Methods), indicated by divergence of regression slopes predicted by a linear mixed effects model accounting for the random effects of ecological region nested in realm and heteroscedasticity (Extended Data Table 2). Both functional diversity and species diversity were calculated using Rao's Q, incorporating abundance information and expressed as effective numbers (reducing species diversity to the inverse Simpson index), with functional diversity also incorporating data on functional traits. Grey shading represents 95% confidence intervals. The Galapagos Islands are distinct from the global pattern.

in the southern hemisphere²³, a bias that has been accentuated in recent years with global focus on declaration of large tropical MPAs²⁴. Our results identify further unrecognized biodiversity value in some temperate and southern hemisphere marine regions, strengthening the argument for greater representation of these areas in global MPA protection. Nevertheless, as is also the case with species diversity, functional trait data need to be coupled with species identity information to ensure, for example, that MPAs within networks are complementary and encompass as much regional biodiversity as possible at both species and functional levels, rather than replicating ecologically similar patches of high diversity only²⁵. Important next steps for building information on functional traits into further conservation applications will include quantification of relationships between metrics of ecological functional and major human impacts (including fishing)²⁶ and development of associated ecological indicators. Widespread use of such indicators offers enormous potential for prioritizing and informing adaptive management and in depicting global conservation targets, alongside indicators tracking decline of rare and endemic taxa²⁷.

METHODS SUMMARY

Field data. Standardized visual censuses along 50-m transects were undertaken at 1,844 sites in 74 of the world's marine ecological regions²⁸ (Extended Data Fig. 2) through the Reef Life Survey program (RLS; <http://www.reeflivesurvey.com>). RLS utilizes skilled volunteer SCUBA divers trained to scientific data collection standards, who undertake surveys with researchers using a consistent methodology that has been evaluated and proven effective for broad-scale studies²⁹. Multiple transects were surveyed at each site (mean 2.4), and species names and abundances of all fishes observed within 5 m either side of transect lines were recorded. Details of quality-control measures are described in the Supplementary Information. Sites were classified as temperate or tropical based on Marine Ecoregions of the World categorization²⁸.

Diversity indices. Species density was the mean number of fish species per 500 m² at each site. Species and functional diversity were calculated using Rao's Q⁶, incorporating relative abundance of each species and converted to effective numbers following a method from a previous study³⁰. Functional diversity and functional

group richness were based on eight functional traits scored for each of the 2,473 fish species in the data set (many from FishBase, <http://www.fishbase.org>; Extended Data Table 1) and calculated using the functional diversity package in R. Evenness was calculated by dividing species diversity by species density (richness).

Global mapping. Relationships between site diversity index values and 13 physical and environmental variables were modelled using random forest methods, where each forest was created by generating 2,000 regression trees from a bootstrap sample of the data. The same technique was used to predict values of each index across 5 arc-minute grid cells globally for mapping. A parallel reduced model was run using only data from sites spaced greater than the distance at which spatial autocorrelation occurred, which confirmed global predictions were not influenced by inherent correlation of diversity values between sites in close proximity.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 May; accepted 6 August 2013.

- Gaston, K. J. Global patterns in biodiversity. *Nature* **405**, 220–227 (2000).
- Hillebrand, H. On the generality of the latitudinal diversity gradient. *Am. Nat.* **163**, 192–211 (2004).
- Rodrigues, A. S. L. *et al.* Effectiveness of the global protected area network in representing species diversity. *Nature* **428**, 640–643 (2004).
- Wiens, J. J. & Donoghue, M. J. Historical biogeography, ecology and species richness. *Trends Ecol. Evol.* **19**, 639–644 (2004).
- Wittebolle, L. *et al.* Initial community evenness favours functionality under selective stress. *Nature* **458**, 623–626 (2009).
- Botta-Dukát, Z. Rao's quadratic entropy as a measure of functional diversity based on multiple traits. *J. Veg. Sci.* **16**, 533–540 (2005).
- Safi, K. *et al.* Understanding global patterns of mammalian functional and phylogenetic diversity. *Phil. Trans. R. Soc. B* **366**, 2536–2544 (2011).
- Stevens, R. D., Cox, S. B., Strauss, R. E. & Willig, M. R. Patterns of functional diversity across an extensive environmental gradient: vertebrate consumers, hidden treatments and latitudinal trends. *Ecol. Lett.* **6**, 1099–1108 (2003).
- Luck, G. W. *et al.* Quantifying the contribution of organisms to the provision of ecosystem services. *Bioscience* **59**, 223–235 (2009).
- Cadotte, M. W. The new diversity: management gains through insights into the functional diversity of communities. *J. Appl. Ecol.* **48**, 1067–1069 (2011).
- Devictor, V. *et al.* Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. *Ecol. Lett.* **13**, 1030–1040 (2010).
- Allen, G. R. Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. *Aquatic Conserv.* **18**, 541–556 (2008).
- Roberts, C. M. *et al.* Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* **295**, 1280–1284 (2002).
- Tittensor, D. P. *et al.* Global patterns and predictors of marine biodiversity across taxa. *Nature* **466**, 1098–1101 (2010).
- Grime, J. P. Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *J. Ecol.* **86**, 902–910 (1998).
- Edgar, G. J., Banks, S., Fariña, J. M., Calvo-piña, M. & Martínez, C. Regional biogeography of shallow reef fish and macro-invertebrate communities in the Galapagos archipelago. *J. Biogeogr.* **31**, 1107–1124 (2004).
- Duffy, J. E. Why biodiversity is important to the functioning of real-world ecosystems. *Front. Ecol. Environ.* **7**, 437–444 (2009).
- Garnfeldt, L. *et al.* Higher levels of multiple ecosystem services are found in forests with more tree species. *Nature Commun.* **4**, 1340 (2013).
- Mora, C. *et al.* Global human footprint on the linkage between biodiversity and ecosystem functioning in reef fishes. *PLoS Biol.* **9**, e1000606 (2011).
- Fonseca, C. R. & Ganade, G. Species functional redundancy, random extinctions and the stability of ecosystems. *J. Ecol.* **89**, 118–125 (2001).
- Mouillot, D. *et al.* Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biol.* **11**, e1001569 (2013).
- Duffy, J. *et al.* Envisioning a marine biodiversity observation network. *Bioscience* **63**, 350–361 (2013).
- Wood, L. J., Fish, L., Laughren, J. & Pauly, D. Assessing progress towards global marine protection targets: shortfalls in information and action. *Oryx* **42**, 340–351 (2008).
- Pala, C. Giant marine reserves pose vast challenges. *Science* **339**, 640–641 (2013).
- Margules, C. R. & Pressey, R. L. Systematic conservation planning. *Nature* **405**, 243–253 (2000).
- Mouillot, D., Graham, N. A. J., Villéger, S., Mason, N. W. H. & Bellwood, D. R. A functional approach reveals community responses to disturbances. *Trends Ecol. Evol.* **28**, 167–177 (2013).
- Jones, J. P. G. *et al.* The why, what, and how of global biodiversity indicators beyond the 2010 target. *Conserv. Biol.* **25**, 450–457 (2011).
- Spalding, M. D. *et al.* Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *Bioscience* **57**, 573–583 (2007).
- Edgar, G. J. & Stuart-Smith, R. D. Ecological effects of marine protected areas on rocky reef communities: a continental-scale analysis. *Mar. Ecol. Prog. Ser.* **388**, 51–62 (2009).
- Jost, L. Entropy and diversity. *Oikos* **113**, 363–375 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the many Reef Life Survey (RLS) divers who participated in data collection and provide ongoing expertise and commitment to the program, University of Tasmania staff including T. Cooper, M. Davey, N. Barrett, J. Berkhout and E. Oh, and T. Bird for assistance running models and checking code. Development of the RLS data set was supported by the former Commonwealth Environment Research Facilities Program, and analyses were supported by the Australian Research Council, Institute for Marine and Antarctic Studies, and the Marine Biodiversity Hub, a collaborative partnership supported through the Australian Government's National Environmental Research Program. Additional funding and support for field surveys was provided by grants from the National Geographic Society, Conservation International, Wildlife Conservation Society Indonesia, The Winston Churchill Memorial Trust, and ASSEMBLE Marine.

Author Contributions R.D.S.-S., J.S.L., G.J.E., S.C.B. and A.E.B. conceived the idea, G.J.E., R.D.S.-S., J.F.S.-S., S.C.B., S.J.K., G.A.S., E.M.A.S. and many others collected the data. R.D.S.-S., A.E.B., J.E.D., G.J.E., and J.S.L. drafted the paper, with substantial input from S.C.B., R.J.T., J.F.S.-S., N.A.H., S.J.K., L.A., M.A.B., S.J.C., T.P.D., S.A.N., G.A.S., E.M.A.S. and T.J.W. A.E.B., R.J.T. and J.S.L. analysed the data and S.J.K. prepared the maps.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Ecological and trait data are available from the authors on request. Correspondence and requests for materials should be addressed to R.D.S.-S. (rstuarts@utas.edu.au).

METHODS

Fish survey methods. Fish abundance data were collected using standard Reef Life Survey (RLS) methods on 4,357 transects at 1,844 sites, in 74 of the world's marine ecological regions (or 'ecoregions', as defined in a previous paper²⁸), with a mean of 2.4 transects per site, at a mean depth of 7.4 m (Extended Data Fig. 2). RLS uses skilled volunteer SCUBA divers trained to scientific data collection standards, who undertake surveys with researchers using a consistent methodology that has been evaluated and proven effective for broad-scale studies^{19,29,31}. This methodology is based on fixed-length belt transects in shallow reef habitats, each covering an area of the seafloor 50×10 m, up to a height of 5 m (full details are provided in an online methods manual; http://reeflifesurvey.com/files/2008/09/NEW-Methods-Manual_15042013.pdf). All fishes sighted in the transect area (by divers swimming either side of the transect) were recorded on an underwater slate, with abundance estimates made by counting individuals of less abundant species and estimating the number of more abundant species. The majority of fishes were recorded to species level, but unidentified fishes were classified at the highest taxonomic resolution possible rather than omitting them. The use of digital photography typically allowed later identification of many unidentified species, with the assistance of taxonomic experts, as required. Appropriate transformations and diversity indices were used to reduce dominating influences of abundance estimates for the most abundant species.

A summary of the total number of transects, species and mean densities of species and individuals in each of the 11 marine realms²⁸ is provided in Extended Data Fig. 2b. This shows clear divisions between tropical and temperate realms in terms of densities of species and individuals, with no temperate realm averaging greater than 18 species and 870 individuals per 500 m², whereas all tropical realms averaged greater than 25 species and 1,430 individuals per 500 m².

The extent and quantity of data used for this study and in the RLS database has only been possible through volunteer efforts. Rigorous measures have been applied to ensure consistency and quality of data, with all divers involved having either substantial prior experience or detailed one-on-one training provided, and extensive data checking applied post dive and before addition to the database. More details on data management and quality control procedures are provided in the Supplementary Information. Only the most skilled subset of divers was involved in the more difficult surveys of diverse coral reefs, where >100 fish species were occasionally recorded on transects.

Functional traits. We compiled values of eight functional traits for each of the 2,473 fish species in the data set (Extended Data Table 1). These were chosen to encompass diverse attributes of fish ecology that are known to influence their functional role in an assemblage, incorporating life history, trophic position, behaviour, and habitat associations. We excluded records of fishes which were not identified to at least Class level (<0.001% of records), and allocated an average value from other members of the same genus (or in some cases, family) for records in which the species was not known, but the genus (or family) was known (0.004% of records). Many trait values were obtained from FishBase (<http://www.fishbase.org/>), and others were allocated based on the authors' combined knowledge of the species.

The allocation of such a comprehensive and informative suite of traits was only possible because fishes are generally so well-known. Despite this, compiling a trait database for reef fishes globally represented an enormous task, and although we attempted to minimize potential errors as much as possible, we acknowledge that a certain level of error is unavoidable. We make the assumption that there is no systematic bias related to the allocation of traits to species from different regions. Further details relating to potential biases in trait allocation are provided in the Supplementary Information.

Diversity indices. Species density was calculated as the number of species observed per 500-m² transect area, whereas species and functional diversity were calculated using Rao's Q⁶, functional group richness (FGR) using a functional dendrogram, functional richness as the convex hull volume, and evenness by dividing species diversity by species density (species richness)³². Rao's Q, FGR and functional richness were calculated using the functional diversity package in R^{33,34} (<http://cran.r-project.org/web/packages/FD/FD.pdf>), incorporating the functional trait data. We excluded transects where fewer than three species were observed before calculating Rao's Q (for functional diversity) and FGR, and transects with fewer than five species for functional richness. A similarity cut-off of 0.18 was used for FGR, based on the maximum length of basal branches in the functional dendrogram using the method 'mcquitty'. Rao's Q (for functional diversity) and functional richness were estimated using the sqrt-corrected species-by-species Gower distance matrix. Functional diversity was subsequently scaled by its maximum value across all frequency distributions to allow for comparisons across sites. Values of Rao's Q (for species and functional diversity) were converted into effective numbers using a linear transformation provided previously³⁰, such that the 'doubling property' was satisfied (also making Rao's Q for species diversity the same as the inverse Simpson index). Functional richness was calculated as a convex hull volume generated from 'new traits' returned following principal coordinates analysis (PCoA) on the corrected

species-by-species distance matrix (R^2 , indicating the quality of the reduced space representation of the traits, was 0.60 with 4 PCoA axes). Gregariousness (1 to 3) was ordered as a factor, and all other traits were coded as unordered factors.

To explore the contributions of individual traits to global patterns in functional diversity, we removed each trait from the full trait matrix and re-calculated Rao's Q (as above) to produce eight functional diversity estimates. We then compared each of the eight functional diversity estimates to functional diversity calculated using the full trait matrix using linear regression. When traits with weaker influence on functional diversity were dropped from the calculation of functional diversity, the change in functional diversity was small and R^2 was close to 1. Those traits that contributed more to functional diversity, when dropped from the trait matrix, resulted in a weaker relationship (lower R^2). This approach therefore allowed the ranking of traits based on the relative change in R^2 when each was dropped from the trait matrix, where lower R^2 values indicated a higher contribution to the overall functional diversity. This process identified water column position, substrate preference and trophic group as contributing most to global patterns in functional diversity (Extended Data Table 3).

The relationship between functional and species diversity presented in Fig. 2 was tested using a linear mixed effects (LME) model, which included the random spatial effects of the ecological region nested in Realm²⁸, and error structures applied to normalize variance in the residuals. We also introduced a spherical autocorrelation structure to our model, however inclusion of this term did not improve the model fit (on the basis of AIC) and this term was therefore removed. Further detail on the model structure and summary information is provided in Extended Data Table 2. The data used were from 1,817 sites (1,844 minus 27 which had fewer than three species recorded – see above) with both species and functional diversity calculated using Rao's Q and expressed as effective numbers (as above). Sites were classified as temperate or tropical based on their location in the Marine Ecoregions of the World categorizations²⁸.

For all methods for quantifying fish communities, visual census methods are characterized by large amounts of variation and a number of biases (for example, see ref. 35). We make the assumption that none of these biases are directional in such a way that will influence the values of diversity metrics over the global scale; for example, that species-specific behavioural traits which affect estimates of their density will be consistent throughout their range (for example, at different latitudes).

Predicting and mapping global diversity measures. Relationships between site diversity metric values and 13 environmental and geographic variables (Extended Data Table 4) were modelled using random forest methods³⁶, in which each forest was created by generating 2,000 regression trees from a bootstrap sample of the data. The same technique was then used to predict values of each index across 5 arc-minute grid cells globally. Environmental variables were obtained from the Bio-ORACLE environmental data set³⁷. A human population index was included as a predictor variable, calculated by fitting a smoothly tapered surface to each settlement point on a year 2000 world-population density grid using a quadratic kernel function described previously³⁸. Populations were screened for a density greater than 1,000 people per 0.04 degree cell, and the search radius was set at 3.959 degrees.

To estimate prediction error, observations not selected in the bootstrap sample for a tree (the 'out-of-bag' sample) were compared to their predictions, in a similar way to cross-validation. The accuracy importance (Extended Data Fig. 3), or mean decrease in accuracy when the predictor variable is randomly permuted, was measured to assess the importance of each predictor variable. Accuracy was evaluated by comparing the predictions with the actual measures of the response variable, for those sites that were out-of-bag for a given tree. Conditional importance of the predictors³⁹ was used, as the standard marginal procedure in the random forest methodology can suffer from bias towards correlated predictors⁴⁰. Biodiversity indices were transformed according to the Box-Cox selection technique (Extended Data Table 5), but all predictions presented have been back-transformed. Random forest modelling was undertaken using the 'extendedForest' packages for R (<https://r-forge.r-project.org/projects/gradientforest>). The fit of each random forest model was measured using the Pearson correlation coefficient between the actual measures and the out-of-bag predictions, before back transformation.

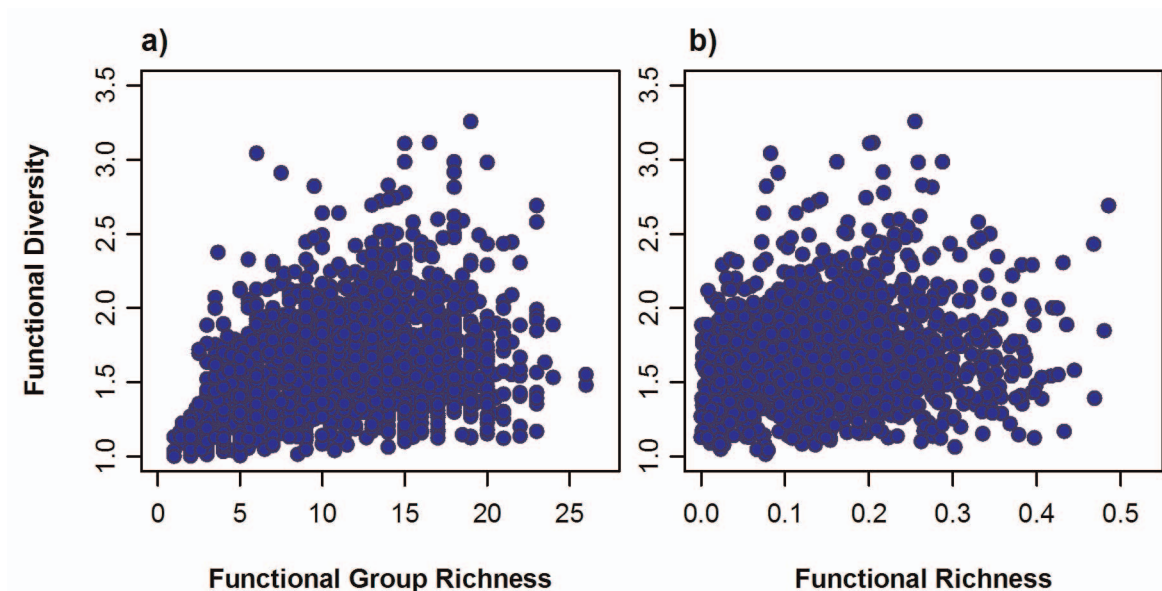
From the generated random forest, predictions were made at new sites by predicting the response variables using each tree individually and then taking the average. Standard deviations of the predictions were used as a measure of prediction error. Predictions of the response variables were made at new sites across the globe and plotted on maps with a coastal buffer.

An assessment of the spatial autocorrelation (SAC) of the global diversity indices was conducted using Moran's I statistic⁴¹, which compared the diversity values at a site with the weighted mean of neighbouring site values. Mean distance of nearest neighbour over all sites was 5.25 km, which was used as the lag distance. All diversity indices were positively spatially autocorrelated in the raw data (Moran's I from 0.23 to 0.61; Extended Data Table 6); an expected result given well-known

patterns in richness are strongly related to latitude and temperature¹⁴, which are also highly spatially autocorrelated. The presence of SAC can typically result in underestimation of model fitting and incorrect hypothesis testing, but model outputs such as mean parameter estimates and predictions are less often affected⁴². There is no hypothesis testing associated with our RF models, and predictions from RF models have previously been identified to be robust to cluster-correlated data, such as in our data set⁴².

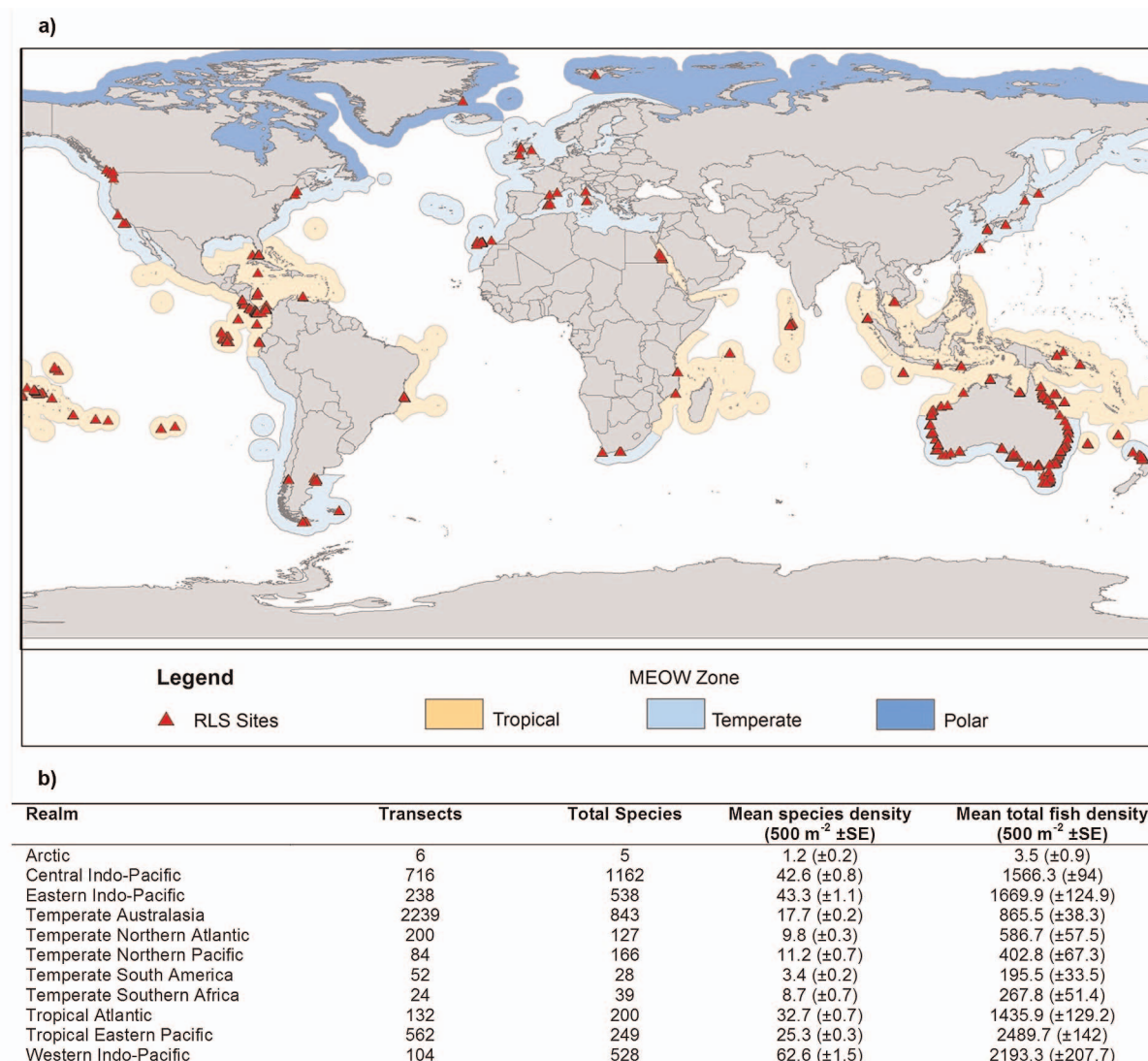
Despite this, we tested for SAC in the residuals of the random forest models to be confident that SAC has not created false patterns in our diversity maps presented in Fig. 1. Although the diversity measures based on the raw data had positive SAC (noted above), the residuals of the random forest models had negative SAC (Extended Data Table 6). Thus, residuals at sites which are closer in distance are more dissimilar than residuals at more distant sites than expected. Visual assessment of variograms showed that the significant negative SAC was at scales less than 5 km for all diversity measures, possibly because the predictor values, and hence the predictions, do not vary greatly within 5 km (owing to the grid size for environmental predictor variables in the Bio-ORACLE data set). Based on this, we generated a second set of random forest models using a subset of sites, chosen with a minimum distance apart of five km (the distance at which SAC became non-significant). Comparison of the original predictions with those made from the subset of sites showed tight concordance (Extended Data Fig. 4), providing strong evidence that diversity patterns presented in our global maps are not unduly influenced by SAC.

31. Edgar, G. J., Barrett, N. S. & Stuart-Smith, R. D. Exploited reefs protected from fishing transform over decades into conservation features otherwise absent from seascapes. *Ecol. Appl.* **19**, 1967–1974 (2009).
32. Tuomisto, H. An updated consumer's guide to evenness and related indices. *Oikos* **121**, 1203–1218 (2012).
33. Laliberté, E. & Legendre, P. A distance-based framework for measuring functional diversity from multiple traits. *Ecology* **91**, 299–305 (2010).
34. Laliberté, E. & Shipley, B. FD: measuring functional diversity from multiple traits, and other tools for functional ecology. R package version 1.0-11 (2011).
35. Edgar, G. J., Barrett, N. S. & Morton, A. J. Biases associated with the use of underwater visual census techniques to quantify the density and size-structure of fish populations. *J. Exp. Mar. Biol. Ecol.* **308**, 269–290 (2004).
36. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
37. Tyberghein, L. *et al.* Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Glob. Ecol. Biogeogr.* **21**, 272–281 (2012).
38. Silverman, B. W. *Density Estimation for Statistics and Data Analysis*. (Chapman and Hall, 1986).
39. Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T. & Zeileis, A. Conditional variable importance for random forests. *BMC Bioinformatics* **9**, 307 (2008).
40. Ellis, N., Smith, S. & Pitcher, C. Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**, 156–168 (2012).
41. Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17–23 10.2307/2332142 (1950).
42. Karpievitch, Y. V., Hill, E. G., Leclerc, A. P., Dabney, A. R. & Almeida, J. S. An introspective comparison of random forest-based classifiers for the analysis of cluster-correlated data by way of RF++. *PLoS ONE* **4**, e7087 (2009).



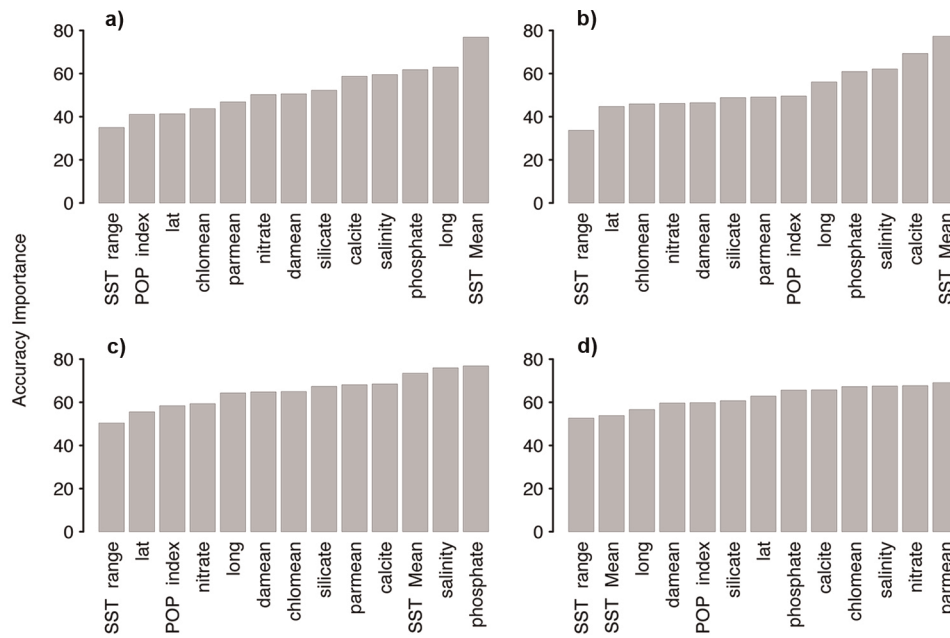
Extended Data Figure 1 | Global relationships between different functional diversity measures. **a, b,** Abundance-weighted functional diversity (Rao's Q expressed as effective numbers; see Methods) for reef fishes provides different information to functional richness expressed as the number of functional

groups (functional group richness; FGR) (**a**) and the volume of multidimensional trait space filled by the community (convex hull volume) (**b**). Kendall's Tau correlation coefficients, $r = 0.21$ and 0.11 , respectively. Points represent individual reef sites surveyed.

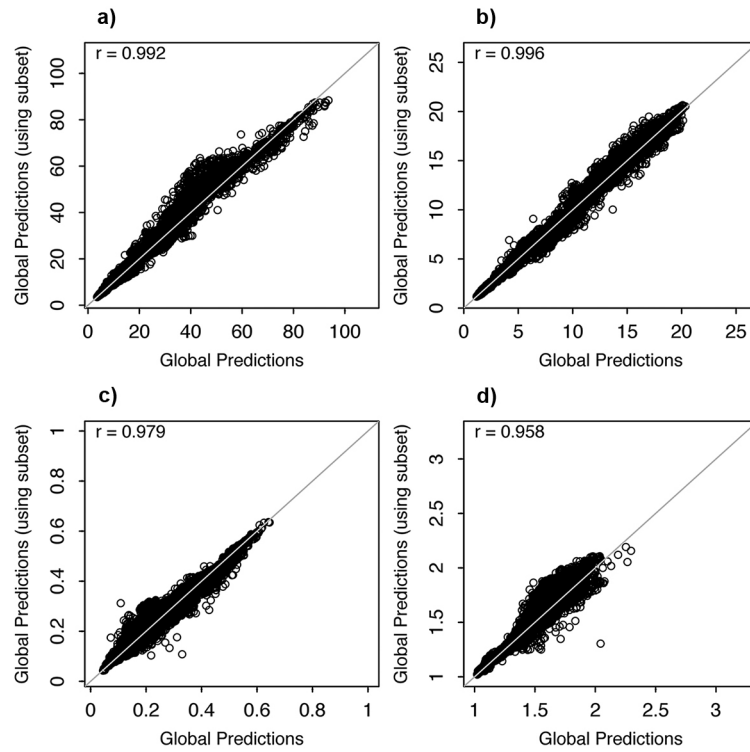


Extended Data Figure 2 | Details of global fish surveys. a, b, The fish fauna was quantitatively surveyed at 1,844 rocky and coral reef sites in 11 Marine Ecoregions of the World realms by visual census (a). Note that many sites are

overlapping or hidden behind symbols for other sites. Tropical realms possessed much higher average fish abundance and species densities (b).



Extended Data Figure 3 | The accuracy importance of the thirteen predictor variables for each of the random forest models. Models were for species density (a), species evenness (b), functional group richness (c) and functional diversity (d). Explanations and units for variables are provided in Extended Data Table 4.



Extended Data Figure 4 | Scatter plots comparing global predictions from random forest models used for mapping in Fig. 1 with those based on a training set including only sites >5 km apart. Models were for species density

(a), functional group richness (b), species evenness (c) and functional diversity (d). Predictions were compared for all global ocean grid cells where the depth was less than 20 m.

Extended Data Table 1 | Functional traits of reef fishes used in estimation of functional diversity.

Functional trait	Category	Type	Units
Maximum length	Body size	Numeric	Total length (cm)*
Trophic breadth	Trophic niche	Numeric	Number of prey phyla consumed (from diet studies*). Range 1-8
Trophic group	Trophic niche	Factor	Browsing herbivore, scraping herbivore, benthic invertivore, planktivore, higher carnivore*
Water column position	Behaviour	Factor	Benthic, demersal, site-attached pelagic, roaming pelagic
Gregariousness	Behaviour	Ordered Factor	Index from 1-3, representing singleton, paired to sometimes forming small schools, always schools
Diel activity pattern	Behaviour	Factor	Nocturnal, diurnal
Preferred substrate	Habitat use	Factor	Hard substrate, soft sediment
Habitat complexity	Habitat use	Factor	Typically associated with habitats characterised by low, medium, high complexity

*Obtained from www.fishbase.org/

Extended Data Table 2 | Linear mixed effects model summary table for functional diversity versus species diversity shown in Fig. 2.

Fixed-effects	Estimate	SE	df	t-value	p-value
Intercept	1.10	0.078	1744	14.05	<0.001
Species Diversity	0.12	0.004	1744	26.91	<0.001
Latitudinal Zone	0.38	0.108	8	3.48	0.008
Species Diversity * Latitudinal Zone	-0.08	0.005	1744	-16.90	<0.001
Random-effects	Realm	Ecoregion	Residual		
	35%	25%	40%		

The percentage variance explained by each random effect ('ecoregion' nested in 'realm') is reported. Error structure (within-realm standard deviation) was included to normalize variance in the residuals. The model was fitted using maximum likelihood with the nlme package in R.

Extended Data Table 3 | Contributions of individual traits to global patterns in functional diversity.

Trait	R^2
Water column position	0.844
Preferred substrate	0.903
Trophic group	0.904
Diel activity pattern	0.908
Habitat complexity	0.934
Gregariousness	0.952
Trophic breadth	0.984
Maximum length	0.998

Traits are listed in order of the magnitude of change in functional diversity arising from their removal, as assessed by R^2 , from the relationship with functional diversity calculated using all traits. Ranks are consistent for other correlation measures.

Extended Data Table 4 | Environmental and geographic variables used in random forest models.

Code	Variable	Unit	Range (reef Sites)
nitrate	Nitrate	$\mu\text{mol/l}$	0.14 - 15.31
SST Mean	Mean Sea Surface Temperature	$^{\circ}\text{C}$	5.35 - 31.09
damean	Diffuse attenuation coefficient at 490 nm	m^{-1}	0.02 - 0.38
phosphate	Phosphate	$\mu\text{mol/l}$	0.02 - 2.26
silicate	Silicate	$\mu\text{mol/l}$	0.31 - 37.9
calcite	Calcite	mol/m^3	$5.2 \cdot 10^{-5}$ - 0.45
chlomean	Mean chlorophyll A concentration	Mg/m^3	0.02 - 15.99
salinity	Salinity	PSS	27.4 - 40.1
parmean	Mean Photosynthetically available radiation	Einstein/ m^2/day	26.3 - 50.9
SST Range	Sea Surface Temperature Range	$^{\circ}\text{C}$	0.96 - 17.93
POP Index	Population Index	People/5 km^2	0 - 1.75 million
lat	Latitude	$^{\circ}\text{N}$	-55.1 - 78.5
long	Longitude	$^{\circ}\text{E}$	-179 - 175

Extended Data Table 5 | Transformations and correlations of observed to predicted diversity values from random forest models.

Diversity metric	Transformation	<i>r</i> (full)	<i>r</i> (reduced)
Species Richness	log	0.93	0.95
Functional Group Richness	Square root	0.91	0.93
Evenness	log	0.67	0.67
Functional Diversity	log	0.58	0.62

Correlations are between recorded values at reef sites and those predicted for the maps in Fig. 1 (full model), and only for sites >5 km apart (reduced).

Extended Data Table 6 | Spatial autocorrelation measured by Moran's I for diversity metrics calculated from the raw data at reef sites and residuals from the four random forest models used to predict diversity values for global maps

	Species richness	Functional group richness	Evenness	Functional diversity
Raw data	0.61*	0.61*	0.31*	0.23*
RF residuals	-0.42*	-0.47*	-0.42*	-0.41*

*denotes significance for two-sided test ($\alpha=0.05$)

A disinhibitory microcircuit initiates critical-period plasticity in the visual cortex

Sandra J. Kuhlman^{1†*}, Nicholas D. Olivas^{2*}, Elaine Tring¹, Taruna Ikrar², Xiangmin Xu^{2,3} & Joshua T. Trachtenberg¹

Early sensory experience instructs the maturation of neural circuitry in the cortex^{1,2}. This has been studied extensively in the primary visual cortex, in which loss of vision to one eye permanently degrades cortical responsiveness to that eye^{3,4}, a phenomenon known as ocular dominance plasticity (ODP). Cortical inhibition mediates this process^{4–6}, but the precise role of specific classes of inhibitory neurons in ODP is controversial. Here we report that evoked firing rates of binocular excitatory neurons in the primary visual cortex immediately drop by half when vision is restricted to one eye, but gradually return to normal over the following twenty-four hours, despite the fact that vision remains restricted to one eye. This restoration of binocular-like excitatory firing rates after monocular deprivation results from a rapid, although transient, reduction in the firing rates of fast-spiking, parvalbumin-positive (PV) interneurons, which in turn can be attributed to a decrease in local excitatory circuit input onto PV interneurons. This reduction in PV-cell-evoked responses after monocular lid suture is restricted to the critical period for ODP and appears to be necessary for subsequent shifts in excitatory ODP. Pharmacologically enhancing inhibition at the time of sight deprivation blocks ODP and, conversely, pharmacogenetic reduction of PV cell firing rates can extend the critical period for ODP. These findings define the microcircuit changes initiating competitive

plasticity during critical periods of cortical development. Moreover, they show that the restoration of evoked firing rates of layer 2/3 pyramidal neurons by PV-specific disinhibition is a key step in the progression of ODP.

Throughout the neocortex, form and function of neural circuitry are shaped by experience^{1,2,7,8}. This sensitivity to experience is most pronounced during adolescence, and has been studied most extensively in the primary visual cortex^{3,9}, in which occluding vision through one eye (monocular deprivation; MD) results in cortical blindness to this eye, even after normal vision is restored^{3,10–12}. With the goal of identifying the initial changes in cortical function after MD we used loose cell-attached recordings to isolate visually evoked responses from single neurons in binocular visual cortex in alert mice (Fig. 1a).

Unexpectedly, we found that visually evoked responses of excitatory pyramidal (PYR) neurons in layer (L)2/3 1 day after contralateral MD were roughly twice as strong as those in control mice. Evoked firing rates to stimulation through the ipsilateral eye, which had not been deprived, were approximately twofold higher than in controls (Fig. 1b, c; MD: 2.9 ± 0.4 Hz, control: 1.5 ± 0.4 Hz; one-way analysis of variance (ANOVA) $P = 0.028$, followed by post-hoc comparison, Tamhane corrected P value $P = 0.046$). Indeed, firing rates evoked by monocular viewing through the open eye after 1 day of MD were sufficiently increased

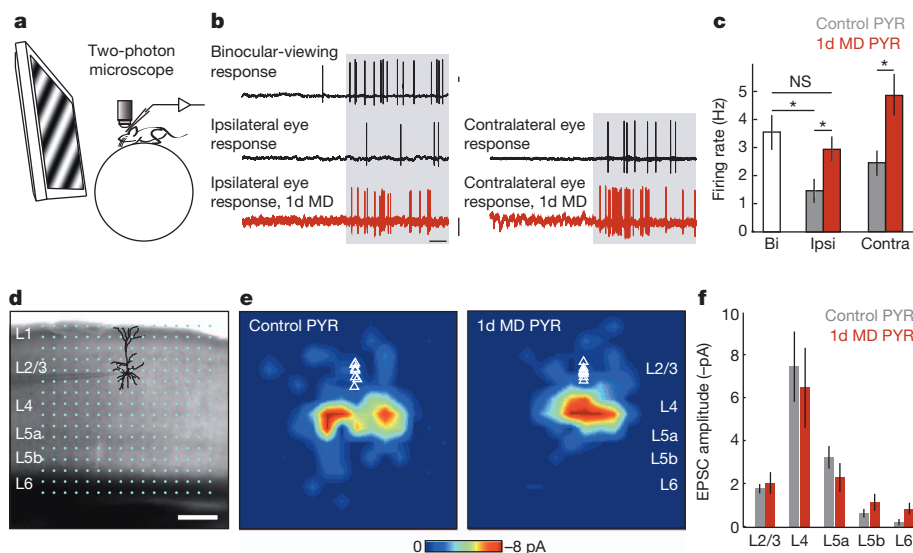


Figure 1 | L2/3 pyramidal neuron responsiveness and local circuit organization is unchanged 1 day after MD. **a–c**, Responses of PYR neurons to drifting gratings in alert mice. **a**, Cartoon of head-fixed configuration. **b**, Example loose cell-attached recordings from controls (black) and after 1-day (1d) MD (red) in response to visual stimulation (grey shading). Scale bar, 1 mV, 500 ms. **c**, Mean firing rate at optimal orientation (binocular (bi): 10 mice, $n = 30$ cells; ipsilateral (ipsi) control: 7 mice, $n = 22$ cells; ipsilateral MD:

6 mice, $n = 33$ cells; contralateral (contra) control: 3 mice, $n = 9$ cells; contralateral MD: 3 mice, $n = 10$ cells). **d**, PYR neuron recorded in binocular V1 in an acute slice; overlaid are 16×16 LSPS stimulation locations spanning pia to white matter. Scale bar, 200 μm . **e**, *In vitro* LSPS aggregate excitatory input maps pooled across PYR neurons. Triangles indicate soma location (control: 4 mice, $n = 9$ cells; MD: 4 mice, $n = 9$ cells). **f**, Mean LSPS-evoked EPSC amplitude, same cells as **e**. NS, not significant. * $P < 0.05$.

¹Department of Neurobiology, David Geffen School of Medicine, University of California, Los Angeles, California 90048, USA. ²Department of Anatomy and Neurobiology, University of California, Irvine, California 92697, USA. ³Department of Biomedical Engineering, University of California, Irvine, California 92697, USA. [†]Present address: Department of Biological Sciences and the Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA.

*These authors contributed equally to this work.

that they were indistinguishable from firing rates evoked during binocular viewing conditions in controls (Fig. 1b, c; binocular: 3.5 ± 0.6 Hz; Tamhane corrected P value $P = 0.834$). A doubling in excitatory firing rates was also seen when we acutely reopened the deprived contralateral eye and recorded responses to visual stimulation while shuttering the ipsilateral eye (Fig. 1b, c; MD: 4.8 ± 0.7 Hz, control: 2.4 ± 0.4 Hz; Mann–Whitney U -test $P = 0.02$). Notably, neurons were not responsive to stimulation through the sutured contralateral eye in alert mice. These findings show that following an immediate drop in visually evoked firing rates after MD (because responses are now only driven through one eye), evoked firing rates of PYR neurons are restored as cortical responsiveness to vision through the open eye increases. Thus, after 1 day of MD the firing rates of PYR neurons in L2/3 of the binocular visual cortex in alert mice that are viewing a scene with only one eye are roughly equivalent to firing rates in normal mice viewing a scene with both eyes.

The increased responsiveness of L2/3 PYR neurons may result from an increase of excitatory drive onto these neurons, enhanced intrinsic excitability, a reduction of inhibitory drive onto these cells, or any combination of these. To examine the possibility that excitatory drive onto L2/3 PYR neurons increases after MD, we assayed the connectivity strength and laminar distribution of presynaptic excitatory inputs onto L2/3 PYR neurons in binocular cortex in acute brain slices using laser scanning photo stimulation (LSPS) via glutamate uncaging¹³ (Fig. 1d–f and Supplementary Fig. 1). In control mice, maps of excitatory laminar inputs to L2/3 PYR neurons were consistent with those shown in previously published studies^{13,14}, with most inputs from L4, L5a and L2/3 as well as some weak input from deeper layers (Fig. 1e, f). MD did not significantly alter the amplitudes of excitatory post-synaptic currents (EPSCs) resulting from stimulation of presynaptic neurons originating in any cortical layer (two-way ANOVA on ranks $P = 0.756$; Fig. 1e, f), indicating that excitatory inputs to L2/3 PYR neurons were not altered at this early stage of plasticity. Spatially restricted photostimulation activates between 50 and 100 neurons at each LSPS location¹⁵, although most of these stimulated neurons are not connected to the neuron being recorded. Because of this, glutamate uncaging may also be used to measure aggregate excitatory synaptic input onto the target neuron by evaluating the number of EPSCs elicited per LSPS location; this measure was also not different between control and 1-day MD (Supplementary Fig. 1). Taken together, local circuit mapping data indicate that 1 day of MD does not significantly change the strength or number of excitatory inputs to L2/3 PYR neurons from any cortical layer. Consistent with a previous report¹⁶, we also did not find any evidence for a significant change in intrinsic excitability of the L2/3 PYR neurons whose inputs we mapped (Supplementary Fig. 1g), nor did we find evidence for a change in resting membrane potential (control: -67.46 ± 1.5 mV, $n = 8$ cells from 3 animals; MD: -65.9 ± 1.4 mV, $n = 7$ cells from 4 animals; Mann–Whitney U -test $P = 0.73$).

A remaining possibility is that a decrease in local inhibition underlies the enhanced visual responsiveness of L2/3 PYR neurons we observe

after MD. Fast-spiking, parvalbumin-positive basket inhibitory neurons, referred to as PV cells, are increasingly implicated in both critical-period and adult plasticity^{5,17}. To measure the impact of 1-day MD on PV responses, we used *in vivo* two-photon imaging to target cell-attached recordings to PV cells expressing the red fluorescent protein tdTomato. Spike waveform analysis was used to verify the identity of recorded neurons (Fig. 2a); recordings were made in alert and urethane-anaesthetized mice. Notably, the effect of 1-day MD on evoked firing rates of PV cells was the opposite of what we found for PYR neurons: whereas PYR spike rates roughly doubled, PV spike rates were roughly reduced by half. In anaesthetized mice, responses to stimulation through the ipsilateral, open eye were reduced by 42% (Fig. 2b, c; control: 16.2 ± 2.0 Hz, MD: 9.5 ± 0.9 Hz; two-way repeated measures ANOVA taking into account the pairwise relationship between ipsilateral and contralateral eye responses recorded for the same cell, $P = 0.002$, followed by post-hoc comparison, Bonferroni-corrected $P = 0.02$), and responses to stimulation through the reopened deprived, contralateral eye were also reduced, in this case by 54% (control: 19.8 ± 2.4 Hz, MD: 9.2 ± 0.9 Hz; Bonferroni-corrected $P = 0.002$). Similarly, in alert mice, PV responses to stimulation through the open eye dropped by 58% relative to controls after 1-day MD (control: 14.8 ± 3.1 Hz, $n = 10$ cells from 8 animals; MD: 6.2 ± 0.9 Hz, $n = 10$ cells from 7 animals; Mann–Whitney U -test $P = 0.001$). These recordings show that MD induces a rapid and severe drop in PV cell firing rates. Importantly, this effect is present for both eyes: responses through either the open or formerly closed eye were reduced by roughly half.

To identify the cause of this rapid, binocular reduction in visually evoked firing rates of PV cells, we again turned to recordings in acute brain slices to measure intrinsic excitability and to map the connectivity strength and laminar position of presynaptic excitatory inputs to PV cells. PV cells expressing tdTomato were visually targeted, their fast-spiking firing patterns were confirmed, and these cells were filled with biocytin for post-hoc morphological analysis. We found no evidence for altered intrinsic excitability as a function of current injection after MD (Supplementary Fig. 2), nor did we detect a change in resting membrane potential (control: -65.4 ± 1.5 mV, $n = 7$ cells from 3 animals; MD: -64.1 ± 0.74 mV, $n = 7$ cells from 4 animals; Mann–Whitney U -test $P > 0.8$). However, glutamate uncaging revealed an $\sim 70\%$ reduction in excitatory drive from cortical layers 4 and upper layer 5 onto L2/3 PV cells after 1-day MD (Fig. 3). The frequency of evoked (Fig. 3b) and spontaneous (Fig. 3c, d) EPSCs were significantly reduced in PV cells after 1-day MD, whereas the amplitude of spontaneous EPSCs was unchanged (Fig. 3e), suggesting both a weakening and reduction of excitatory synaptic inputs onto PV cells occurring at these earliest stages of cortical plasticity.

Notably, the rapid binocular reduction of PV firing rates after MD is transient. When MD was extended to 3 days, PV responses to stimulation of the deprived, contralateral eye (measured after acutely opening the sutured lids) remained as weak as they were after 1-day MD, but

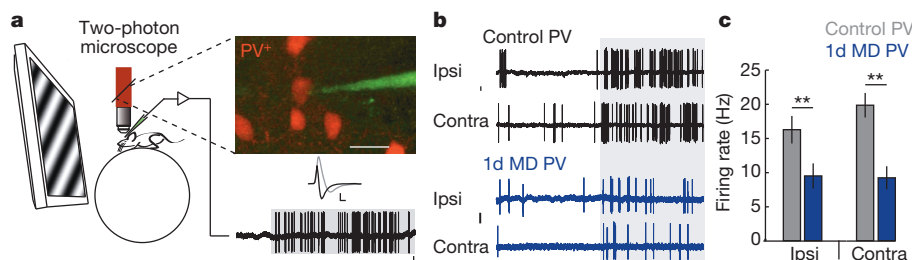


Figure 2 | L2/3 PV responsiveness to visual stimuli is reduced after 1-day MD. **a**, Cartoon of targeted recording in alert mice (left). Two-photon excitation (red beam) is used to visualize PV cells expressing tdTomato and recording pipette filled with Alexa dye (inset, 6 red PV cells; the pipette (green) is loose cell-attached to the PV cell in the centre of the image; scale bar, 20 μ m). Spike waveform is used to verify targeting of fast-spiking PV cells (black trace; grey trace is a PYR neuron waveform for comparison; scale bar, 0.5 mV, 1 ms).

Bottom, example PV response to visual stimulation (grey shading; scale bar, 1 mV, 500 ms). **b**, Example PV responses evoked by stimulation through either eye in control (black) and 1-day MD (blue) anaesthetized mice. Scale bar, 1 mV, 500 ms. **c**, Mean firing rate at optimal orientation (ipsilateral control: 14 mice, $n = 26$ cells; ipsilateral MD: 5 mice, $n = 18$ cells; contralateral control: 14 mice, $n = 26$ cells; contralateral MD: 5 mice, $n = 18$ cells). $^{**}P < 0.005$.

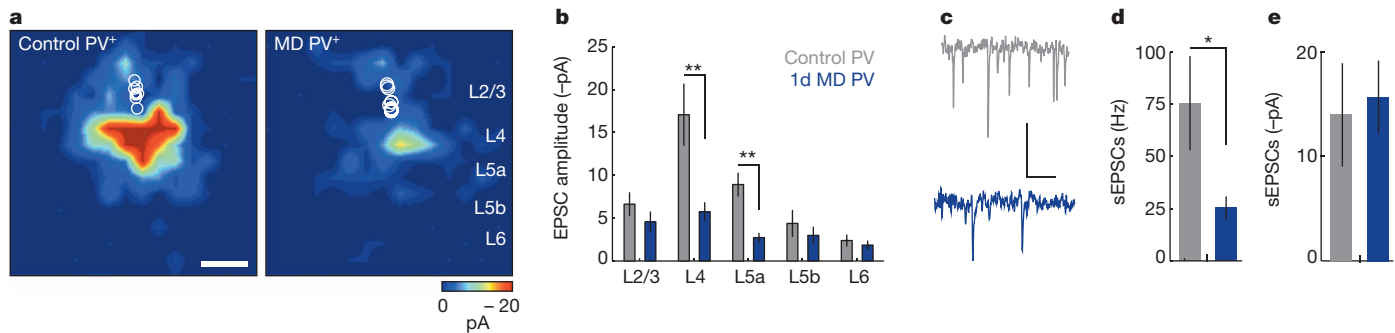


Figure 3 | L2/3 PV local circuit organization is altered after 1-day MD. **a**, *In vitro* LSPS aggregate excitatory input maps pooled across PV cells. Circles indicate soma location (control: 4 mice, $n = 7$ cells; MD: 4 mice, $n = 7$ cells). Scale bar, 200 μm . **b**, Mean laminar LSPS-evoked EPSC amplitude, same cells as

a. $**P < 0.007$. **c**, Example spontaneous EPSCs (sEPSCs). Scale bar, 20 pA, 100 ms. **d**, Mean sEPSC frequencies, same cells as **a**. $*P < 0.05$. **e**, Mean sEPSC amplitude, same cells as **a**.

the strength of PV responses evoked by ipsilateral eye stimulation returned to normal (Supplementary Fig. 3). Thus, the rapid reduction in PV spike rates after 1-day MD is soon replaced by a restoration of normal PV responses to ipsilateral eye stimulation, which shifts PV ocular dominance towards the open eye.

A key question emerging from these findings is whether the transient binocular disinhibition following lid suture is necessary for subsequent shifts in excitatory ocular dominance. If so, one prediction is that maintaining inhibition after MD will prevent excitatory plasticity. To examine this we enhanced inhibition by injecting the use-dependent GABA_A (γ -aminobutyric acid receptor-A) agonist diazepam into the ventricles of postnatal day (P)28 mice immediately after suturing the contralateral eye. Diazepam treatment effectively prevented the increase in ipsilateral eye excitatory firing rates seen after 1-day MD, but did not silence evoked activity (Supplementary Fig. 4a, b).

Cortical ocular dominance also was assayed 3 days after MD by imaging visually evoked fluorescent changes in L2/3 neurons bulk loaded with the calcium indicator Oregon Green BAPTA-1. Mice were exposed to 3 days of MD, and received intraventricular injections of either saline or diazepam for the first 2 days. Consistent with a role for disinhibition in ODP, intraventricular diazepam, but not vehicle, prevented shifts in cortical ocularity after MD (Supplementary Fig. 4c–h).

Because ODP is restricted to the critical period⁹, a second prediction is that MD will not alter PV firing rates when performed thereafter. Supporting this we found no significant changes in evoked PV firing rates in monocularly deprived adult mice (age P42–45) relative to age-matched controls (Supplementary Fig. 5).

A third prediction is that ODP will be enhanced in older mice if PV spike rates are artificially reduced after MD. There is some precedent for this view¹⁸. In order to manipulate PV spike rates *in vivo*, we expressed hM₄D-inactivating G_i-coupled DREADD receptors^{19–22} selectively in PV cells. Binding of the ligand clozapine N-oxide (CNO) to hM₄D receptors activates G-protein-coupled inwardly rectifying potassium channels, thereby reducing spike rates in target neurons for 9–12 h²³ (Supplementary Fig. 6). Three days of MD beginning at P35 does not normally shift cortical ocular dominance⁹. To determine whether a transient reduction in PV-mediated inhibition could reveal ODP in mice of this age, we tracked changes in cortical ocular dominance by imaging cortical responses to ipsilateral and contralateral eye stimulation longitudinally in mice expressing the genetically encoded calcium indicator GCaMP6 (ref. 24) in L2/3 of the binocular visual cortex. P35 mice given two injections of CNO at 12-h intervals immediately after lid suture (1 day of reduced PV inhibition) showed significant decreases in cortical responsiveness to the deprived eye without any significant

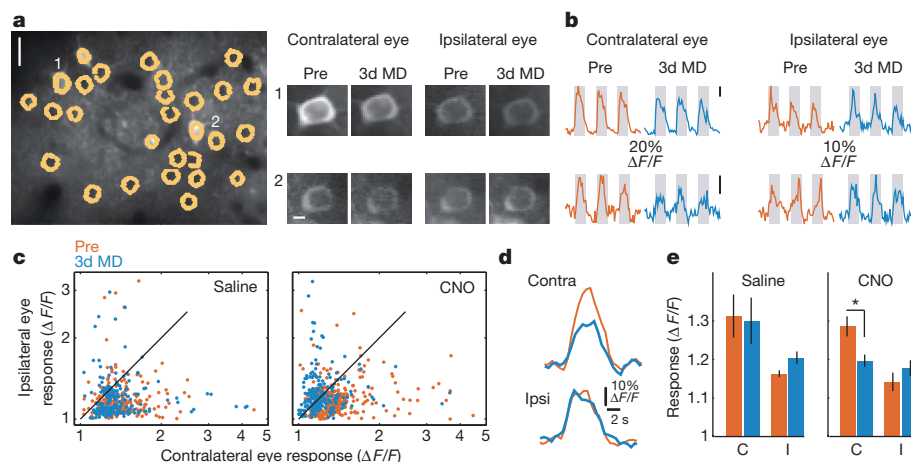


Figure 4 | Reducing PV-specific inhibition restores ODP after the closure of the critical period. **a**, Left, *In vivo* image of GCaMP6 expression, cells outlined in yellow. Scale bar, 20 μm . Right, individual images for cells labelled 1 and 2 of the average evoked fluorescence in response to visual stimuli presented independently to the contralateral and ipsilateral eye, before (pre) and 3 days after (3d) MD in a mouse expressing the hM₄D DREADD receptor specifically in PV cells, treated with CNO. Scale bar, 5 μm . **b**, Relative change in fluorescence before and after 3-day MD + CNO for cells 1 and 2 in panel **a** in response to visual stimulation (grey shading; 5-s duration). **c**, Log-log scatter plot of visually evoked fluorescence response for each cell (saline: $n = 242$ cells;

CNO: $n = 327$ cells). Note the significant leftward shift after 3-day MD in mice treated with CNO (2-way ANOVA interaction, $P = 0.037$), but not those treated with saline (2-way ANOVA interaction, $P = 0.72$), indicating a reduced response to contralateral eye stimulation after 3-day MD. **d**, Longitudinal 'optical field potential' response (average value of all pixels in the entire imaging field, inclusive of neuropil) for the region in **a**. Note the decrease in response to contralateral eye stimulation after MD + CNO that is not seen for the ipsilateral eye. **e**, Mean 'optical field potential' response (saline: $n = 4$ mice; CNO: $n = 4$ mice). $*P < 0.05$.

change in open, ipsilateral eye responses, whereas age-matched controls receiving similar injections of saline showed no changes in either measure (Fig. 4). Thus, a 24-h period of diminished PV firing is sufficient to restore juvenile-like ODP after the closure of the critical period.

The shift in ocular dominance after eye closure has been a long-standing paradigm for studying cortical plasticity. Mounting evidence implicates PV inhibitory neurons as regulators of this plasticity. How they do so, however, has remained incompletely understood. The data we present here define the role of PV cells in this process by showing that MD results in a rapid, but transient, reduction in PV-based inhibition, which restores PYR firing rates to pre-deprivation levels. Without this restoration of normal excitatory firing rates, subsequent competitive plasticity will not occur.

Our data and earlier papers examining the plasticity of the GABA population as a whole^{25,26} or specifically PV cells²⁷ report that inhibitory responses ultimately shift towards the open eye after MD. Furthermore, there is a consensus that in non-deprived conditions, inhibitory cells are more binocular than excitatory cells in mice with normal GABA levels^{25,27}. However, our data differ considerably from earlier studies on two fundamental issues: (1) the rate and direction of PV cell plasticity; and (2) the role of PV cell plasticity in excitatory ODP. Specifically, the possibility that PV cells become transiently, and paradoxically more responsive to deprived eye stimulation, thereby suppressing deprived-eye responses of excitatory neurons and shifting excitatory ocular dominance towards the open eye, was raised in a study that measured the ratio of the contralateral and ipsilateral eye responses in PV cells after brief MD²⁷. In our measures of PV spike rates we did not find this paradoxical shift. Instead, we found that a rapid, transient and binocular reduction of fast-spiking inhibition permits, but does not instruct, synaptic competition among excitatory neurons. Our data redefine the role of fast-spiking inhibitory neurons as gating excitatory plasticity. It is the reduction of inhibition and the restoration of normal excitatory spike rates that permits competitive ODP to proceed on the basis of the relative strengths of the inputs of the two eyes.

Distinct mechanisms govern the influence of sensory experience on cortical plasticity early in life, in the form of passive sensory exposure, and in adulthood, in the form of reinforced associative learning²⁸. The earliest stages of reinforced associative learning in primary sensory cortices appear to involve an active inhibition of PV inhibitory neurons via cholinergic activation of layer 1 interneurons^{17,29}. Notably, our results indicate that a similar strategy—a reduction in PV cell firing rates and disinhibition of upper layer excitatory neurons—is used during the initial stages of binocular plasticity in adolescence, although by a loss of excitatory drive onto the PV cells. Together, these observations indicate that pathways regulating cortical plasticity in the juvenile and adult cortex converge on a rapid reduction of PV cell activity.

METHODS SUMMARY

All *in vivo* recordings were conducted in accordance with procedures approved by the University of California Los Angeles Office for Protection of Research Subjects and the Chancellor's Animal Research Committee. For work using cortical slices, the Institutional Animal Care and Use Committee at the University of California, Irvine, approved all procedures.

Full Methods and any associated references are available in the online version of the paper.

Received 8 October 2012; accepted 17 July 2013.

Published online 25 August 2013.

1. Ko, H. *et al.* The emergence of functional microcircuits in visual cortex. *Nature* **496**, 96–100 (2013).
2. White, L. E. & Fitzpatrick, D. Vision and cortical map development. *Neuron* **56**, 327–338 (2007).
3. Wiesel, T. N. & Hubel, D. H. Single-cell responses in striate cortex of kittens deprived of vision in one eye. *J. Neurophysiol.* **26**, 1003–1017 (1963).

4. Levelt, C. N. & Hubener, M. Critical-period plasticity in the visual cortex. *Annu. Rev. Neurosci.* **35**, 309–330 (2012).
5. Hensch, T. K. Critical period plasticity in local cortical circuits. *Nature Rev. Neurosci.* **6**, 877–888 (2005).
6. Jiang, B., Huang, Z. J., Morales, B. & Kirkwood, A. Maturation of GABAergic transmission and the timing of plasticity in visual cortex. *Brain Res. Brain Res. Rev.* **50**, 126–133 (2005).
7. Fu, M. & Zuo, Y. Experience-dependent structural plasticity in the cortex. *Trends Neurosci.* **34**, 177–187 (2011).
8. Holtmaat, A. & Svoboda, K. Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Rev. Neurosci.* **10**, 647–658 (2009).
9. Gordon, J. A. & Stryker, M. P. Experience-dependent plasticity of binocular responses in the primary visual cortex of the mouse. *J. Neurosci.* **16**, 3274–3286 (1996).
10. Webber, A. L. & Wood, J. Amblyopia: prevalence, natural history, functional effects and treatment. *Clin. Exp. Optom.* **88**, 365–375 (2005).
11. Holmes, J. M. *et al.* Effect of age on response to amblyopia treatment in children. *Arch. Ophthalmol.* **129**, 1451–1457 (2011).
12. Kanonidou, E. Amblyopia: a mini review of the literature. *Int. Ophthalmol.* **31**, 249–256 (2011).
13. Shepherd, G. M. & Svoboda, K. Laminar and columnar organization of ascending excitatory projections to layer 2/3 pyramidal neurons in rat barrel cortex. *J. Neurosci.* **25**, 5670–5679 (2005).
14. Yoshimura, Y. & Callaway, E. M. Fine-scale specificity of cortical networks depends on inhibitory cell type and connectivity. *Nature Neurosci.* **8**, 1552–1559 (2005).
15. Weiler, N., Wood, L., Yu, J., Solla, S. A. & Shepherd, G. M. Top-down laminar organization of the excitatory network in motor cortex. *Nature Neurosci.* **11**, 360–366 (2008).
16. Lambo, M. E. & Turrigiano, G. G. Synaptic and intrinsic homeostatic mechanisms cooperate to increase L2/3 pyramidal neuron excitability during a late phase of critical period plasticity. *J. Neurosci.* **33**, 8810–8819 (2013).
17. Letzkus, J. J. *et al.* A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature* **480**, 331–335 (2011).
18. Harauzov, A. *et al.* Reducing intracortical inhibition in the adult visual cortex promotes ocular dominance plasticity. *J. Neurosci.* **30**, 361–371 (2010).
19. Ferguson, S. M. *et al.* Transient neuronal inhibition reveals opposing roles of indirect and direct pathways in sensitization. *Nature Neurosci.* **14**, 22–24 (2011).
20. Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S. & Roth, B. L. Evolving the lock to fit the key to create a family of G protein-coupled receptors potentially activated by an inert ligand. *Proc. Natl Acad. Sci. USA* **104**, 5163–5168 (2007).
21. Dong, S., Allen, J. A., Farrell, M. & Roth, B. L. A chemical-genetic approach for precise spatio-temporal control of cellular signaling. *Mol. Biosyst.* **6**, 1376–1380 (2010).
22. Li, H. *et al.* Experience-dependent modification of a central amygdala fear circuit. *Nature Neurosci.* **16**, 332–339 (2013).
23. Alexander, G. M. *et al.* Remote control of neuronal activity in transgenic mice expressing evolved G protein-coupled receptors. *Neuron* **63**, 27–39 (2009).
24. Chen, T.-W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
25. Kameyama, K. *et al.* Difference in binocularity and ocular dominance plasticity between GABAergic and excitatory cortical neurons. *J. Neurosci.* **30**, 1551–1559 (2010).
26. Gandhi, S. P., Yanagawa, Y. & Stryker, M. P. Delayed plasticity of inhibitory neurons in developing visual cortex. *Proc. Natl Acad. Sci. USA* **105**, 16797–16802 (2008).
27. Yazaki-Sugiyama, Y., Kang, S., Cateau, H., Fukai, T. & Hensch, T. K. Bidirectional plasticity in fast-spiking GABA circuits by visual experience. *Nature* **462**, 218–221 (2009).
28. de Villers-Sidani, E. & Merzenich, M. M. Lifelong plasticity in the rat auditory cortex: basic mechanisms and role of sensory experience. *Prog. Brain Res.* **191**, 119–131 (2011).
29. Froemke, R. C., Merzenich, M. M. & Schreiner, C. E. A synaptic memory trace for cortical receptive field plasticity. *Nature* **450**, 425–429 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Smith, W. Thompson, K. Miller and M. P. Stryker for comments on earlier versions of this manuscript, Y. Shi for help with software, Z. Nenadic for analytical suggestions, D. Ringach for help with GCaMP6 analysis and Z. J. Huang for useful discussions. This work was funded by grants from the US National Eye Institute (EY016052) to J.T.T., and the US National Institute of Neurological Disorders and Stroke (NS078434) and a NARSAD Young Investigator Grant to X.X.

Author Contributions S.J.K. and E.T. performed the *in vivo* awake and anaesthetized recordings. S.J.K. performed the diazepam/OGB-1 experiments. E.T. performed the DREADD/GCaMP6 experiments. N.D.O. and T.I. performed the glutamate uncaging experiments. X.X. oversaw the glutamate uncaging experiments. J.T.T. oversaw all aspects of the project. S.J.K., N.D.O., X.X. and J.T.T. wrote the manuscript and prepared the figures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.X. (xiangmix@uci.edu) or J.T.T. (joshua.trachtenberg@gmail.com).

METHODS

All *in vivo* recordings were conducted in accordance with procedures approved by the University of California Los Angeles Office for Protection of Research Subjects and the Chancellor's Animal Research Committee. For work using cortical slices, the Institutional Animal Care and Use Committee at the University of California, Irvine, approved all procedures. To genetically label PV neurons, PV-IRES-cre knock-in female mice (Jackson Laboratories, stock no. 008069, generated by S. Arbor, FMI) were crossed with male tdTomato reporter knock-in mice directly received from Jackson Laboratory (Jackson Laboratories, stock no. 007905, 'Ai9', generated by H. Zeng, Allen Brain Institute). All experimental mice were hemizygous for both transgenes (PV-Cre:Ai9). Homozygous PV-IRES-cre mice used for the above breeding were all from a F₁ cross of a male and female directly received from Jackson Laboratory. Mice (male and female) were randomly assigned to either control or MD treatment groups. Sample sizes were chosen using a target power of 0.8. For cases in which analysis included a subjective component, analysis was done blind to experimental group as noted below. MD was performed under isoflurane/oxygen anaesthesia (3% induction and 2% maintenance isoflurane), antibacterial ointment applied to eyes, lid margins trimmed and one mattress suture (silk 6-0) was used to bind the upper and lower lids. In the case of 1-day MD alert recording experiments, lid margins were not trimmed to facilitate reopening. Any mice showing signs of eye infection or lid separation were removed from the study.

Electrophysiological recordings in alert mice. We followed the protocol outlined in ref. 30. To habituate mice to head restraint, mice were handled on the first day of training by repeatedly picking them up. On the second day, a metal bar later used to restrain the animal was fixed to the mouse's skull using dental acrylic. After a minimum of 2 days recovery, mice were placed on a floating styrofoam ball, head fixed and allowed to run for two 10-min sessions with a 10-min break in between. This was repeated daily, without the 10-min break, until mice showed signs of controlled navigation, including smooth transitions between running, walking, grooming and balancing without motion. Typically 2 days of head restraint atop the spherical treadmill was sufficient. On the day of recording, a 2-mm diameter craniotomy was made directly over the binocular zone of the primary visual cortex. Craniotomy placement and *in vivo* electrophysiology were performed as detailed in ref. 31. Craniotomies were centred 3 mm lateral of the sagittal suture and 1 mm rostral of the lambdoid suture. All experimental animals contained neurons that were driven by both the ipsilateral and contralateral eye in the recording zone. We refer to all non-fast spiking neurons (narrow-width action potentials²) as excitatory pyramidal neurons, although we do not rule out the possibility that up to 10% of the neurons are inhibitory neurons. Treadmill motion was monitored by measuring intensity changes in the reflectance of a red laser beam (Keyence Corporation, part no. LV-N11MN and LV-NH32) focused on the styrofoam ball. Electrophysiological signals were acquired in cell-attached configuration using a Multiclamp 700B amplifier (Molecular Devices) in current-clamp mode, digitized (PCI-6229M, National Instruments) at a sample rate of 10 kHz and low-pass filtered at 6 kHz. This method of recording is ideally suited to measuring changes in firing rates because there is not an inherent sampling bias for more active neurons, a bias that is present in other techniques that depend on isolating single units from multiunit activity.

Electrophysiological recordings in urethane anaesthetized mice. We followed the protocol outlined in ref. 31.

Visual stimulation and analysis of *in vivo* electrophysiology. To measure evoked firing rate at the preferred orientation, full-field drifting square wave gratings were presented at full contrast at 6 orientations spaced 30° apart at two directions of motion (12 orientations), using custom software developed with Psychtoolbox in Matlab. A 40-cm-width LCD monitor was placed 25 cm directly in front of alert or urethane-anaesthetized mice³¹, tangent to the animal's gaze. Each stimulus was presented for 3 s, followed by a grey screen of equal duration. Stimulus orientations and direction were randomized. Varying spatial frequencies (0.02 and 0.04 cycles per degree) and temporal frequencies (1–2 Hz) were tested, and optimal stimulation was then shown for 4–9 trials. In awake recordings, periods of grooming and any periods >60 s in which there was no treadmill motion detected were excluded from further analysis. Evoked rates were computed as the spike rate for a given orientation, averaged across the 3-s stimulation window and across 4–9 trials, and were baseline-subtracted. Baseline was defined as the average spike rate in response to all grey screen presentations. Evoked rate at the preferred orientation is reported. To analyse spike shape, the first 50 spike waveforms were averaged for a given cell. The 10–90% rising and falling slope, and peak (P1) and nadir (P2) amplitudes were calculated from the average trace.

Eye shuttering. Eye shuttering was accomplished by manually placing an occluding device 5 mm in front of the eye. The occluding device was constructed of flexible light-blocking material (1.2 × 1.5 cm) mounted to a vertical post, 5 mm in diameter, coupled to a magnetic base that was released and alternately placed in front of either eye as indicated by the interleaved trial schedule.

Visual cortical slice preparations. The littermates of double-transgenic mice (PV-Cre:Ai9, aged P28–P29) were used for either control or the treatment of 24-h MD. For each experimental day, a pair of the control and MD-treated animals was deeply anaesthetized using pentobarbital sodium (100 mg kg⁻¹, intraperitoneally), transcardially perfused with chilled oxygenated artificial cerebrospinal fluid (ACSF), rapidly decapitated and brain extracted. Coronal sections of 400 µm were cut from primary visual cortex (V1) with a vibratome (VT1200S, Leica Systems) in sucrose containing ACSF (in mM: 85 NaCl, 75 sucrose, 2.5 KCl, 25 glucose, 1.25 NaH₂PO₄, 4 MgCl₂, 0.5 CaCl₂ and 24 NaHCO₃). For animals having undergone MD, sections were taken exclusively from the hemisphere contralateral to the deprived eye. Slices were incubated for at least 30 min in sucrose containing ACSF at 32 °C before being transferred into slice recording chambers with normal ACSF (in mM: 126 NaCl, 2.5 KCl, 26 NaHCO₃, 2 CaCl₂, 2 MgCl₂, 1.25 NaH₂PO₄ and 10 glucose). Throughout the cutting, incubation and recording, the solutions were continuously supplied with 95% O₂–5% CO₂.

Electrophysiology and LSPS. Electrophysiological recordings and photostimulation were performed as in ref. 32. Whole-cell recordings were performed under a DIC/fluorescent Olympus microscope (BX51WI). Oxygenated ACSF at room temperature (25 °C) was perfused into the slice recording chamber through a custom-designed flow system driven by pressurized 95% O₂–5% CO₂ (3 PSI) at roughly 2 ml min⁻¹. Slices were examined under a 4× objective for proper targeting of either L2/3 pyramidal neurons or tdTomato-expressing PV-positive interneurons within binocular regions of mouse V1 using landmarks defined in ref. 33. To target whole-cell recordings, cells were visualized at high magnification (60× objective, 0.9 NA; LUMPlanFI/IR, Olympus). Cell bodies of recorded neurons were at least 50 µm below the surface of the slice. Patch pipettes (4–6 MΩ resistance) made of borosilicate glass were filled with an internal solution containing (in mM) 126 K-glucuronate, 4 KCl, 10 HEPES, 4 ATP-Mg, 0.3 GTP-Na and 10 phosphocreatine (pH 7.2, 300 mOsm). Electrodes also contained 0.1% biocytin for post-hoc cell labelling and further morphological identification. Once stable whole-cell recordings were achieved with good access resistance (usually <20 MΩ), basic electrophysiological properties were examined through hyperpolarizing and depolarizing current injections. Electrophysiological data were acquired with a Multiclamp 700B amplifier (Molecular Devices), data acquisition boards (models PCI MIO 16E-4 and 6713, National Instruments) and custom-modified version of Ephus software³⁴ (Ephus, available at <https://openwiki.janelia.org/>). Data were digitized at 10 kHz.

PV-positive inhibitory neurons were targeted on the basis of RFP expression and verified by fast-spiking patterns from current injections and large basket cell morphology. Excitatory neurons were selected on the basis of their pyramidal somata detected under DIC, were RFP negative and showed adapting spiking patterns to suprathreshold intrasomatic current injections. Final cell type classification was determined by the combined characterization of RFP expression, electrophysiological and morphological properties of the recorded cells.

The LSPS procedures were similar to those described in refs 13 and 35. LSPS was performed through a 4× objective lens. Stock solution of MNI-caged-L-glutamate (Tocris Bioscience) was added to 20 ml ACSF for a concentration of 0.2 mM caged glutamate. The cortical slice image, acquired through the 4× objective, was visualized using a high-resolution digital CCD camera, and this image, in turn, was used to guide and register photostimulation sites. Two-millisecond duration, 20-mW pulses from a 350 nm UV laser (DPSS Lasers) were delivered to the sample, controlled via an electro-optical modulator and a mechanical shutter. Focal laser spots approximated a Gaussian profile with a lateral width of 100 µm. Under our experimental conditions, LSPS-evoked action potentials were recorded from stimulation locations within 124 µm (96 ± 28 µm, *n* = 12) of targeted somata and occurred within 150 ms after photostimulation. Synaptic currents in patched neurons were detected under voltage clamp. By systematically surveying synaptic inputs from hundreds of different sites across a large cortical region, aggregate synaptic input maps were generated for individual neurons. For our mapping experiments, a standard stimulus grid (16 × 16 stimulation sites, 65 µm² spacing) was used to tessellate V1 from pia to white matter. The LSPS site spacing was empirically determined to capture the smallest predicted distance in which photostimulation differentially activates adjacent neurons. Glutamate uncaging was delivered sequentially in a nonraster, nonrandom sequence, following a 'shifting-X' pattern designed to avoid revisiting the vicinity of recently stimulated sites³⁶.

Because glutamate uncaging activates both excitatory and inhibitory neurons, we empirically determined the excitatory and inhibitory reversal potentials in L2/3 pyramidal cells to properly isolate EPSCs and IPSCs. In a subset of experiments, L2/3 neurons were voltage clamped at a range of holding potentials from –90 mV to 5 mV, and GABA uncaging (0.1 mM, Invitrogen) was performed by delivering brief ultraviolet flashes (1 ms, 20 mW) to the perisomatic regions of targeted cells to determine the GABAergic reversal potential on the basis of voltage–current relationships. The average GABAergic reversal potential ranged from –69 mV to

−71 mV, and we therefore voltage clamped the targeted pyramidal cells at −70 mV to determine LSPS and spontaneously evoked EPSCs.

Laminar circuit input analysis. Photostimulation induces two forms of excitatory responses: (1) those that result from direct activation of the recorded neuron's glutamate receptors; and (2) synaptically mediated responses (EPSCs) resulting from the suprathreshold activation of presynaptic excitatory neurons. Responses that occur within 10 ms of laser pulse onset were considered direct; these responses exhibited a distinct shape and occurred immediately after glutamate uncaging. Synaptic currents with such short latencies are not possible because they would have to occur before the generation of action potentials in photostimulated neurons. Therefore, direct responses need to be excluded from local synaptic input analysis. To check for any systematic differences across treatment conditions, the spatial extent and frequency of action potentials elicited in response to direct stimulation was determined in a subset of the experiments by performing whole-cell recordings in current-clamp mode using an 8×8 mapping grid. Direct stimulation was found to be similar for control and MD; direct responses were restricted to within approximately 100 μm of the recorded soma (control: $86 \pm 22 \mu\text{m}$, $n = 7$ cells; MD: $106 \pm 34 \mu\text{m}$, $n = 5$ cells) and were limited to less than 1.5 action potentials averaged across trials (control: 1 ± 0 , $n = 7$ cells; MD: 1.2 ± 0.2 , $n = 5$ cells). At some locations, synaptic responses were over-riding on relatively small direct responses (see Supplementary Fig. 1d); such responses were identified and included in synaptic input analysis.

For data map analysis, we implemented the approach for detection and extraction of photostimulation-evoked postsynaptic current responses described in ref. 37. LSPS-evoked EPSCs were quantified across the 16×16 mapping grid for each cell, and 2 to 4 individual maps were averaged per recorded cell, reducing the likelihood of incorporating noise events in the analysis window (150 ms). Averaged maps were then analysed using the $4 \times$ DIC image to bin responses according to laminar cytoarchitectonic landmarks. Synaptic events were binned from locations spanning $\pm 195 \mu\text{m}$ tangential to the targeted soma location and from the top of layer 2/3 to the bottom of layer 6 across the radial vector. Data were plotted as either the average integrated EPSCs amplitude per pixel location or the number of EPSCs detected per pixel location.

Morphological examination and cell-type Identification. The morphology of each recorded neuron was determined using post-hoc staining. In brief, brain slices were fixed in 4% paraformaldehyde and transferred to 30% sucrose solution in PBS. Neurons filled with biocytin during recordings were labelled with Alexa Fluor 488-conjugated streptavidin (1:500 dilution; Jackson ImmunoResearch). Slices were also stained for 4'-6-diamidino-2-phenylindole (DAPI) (Sigma-Aldrich) to identify laminar boundaries. Cell morphology, DAPI labelling and RFP expression were visualized using an Olympus BX 61 epifluorescent microscope and MetaMorph imaging suite (Molecular Devices).

Calcium imaging using Oregon Green BAPTA-1. Mice were anaesthetized with urethane and core body temperature was maintained using a closed loop heated plate and a rectal temperature sensor. Neurons in L2/3 of the primary visual cortex were bulk labelled with the calcium indicator Oregon Green BAPTA-1 (OGB), and changes in fluorescence intensity were imaged as described in refs 38 and 39. Dye was injected at 2–3 positions to ensure labelling within the binocular region. Three days before calcium imaging, mice were randomly assigned to one of three groups: control, MD or MD plus diazepam (pharmaceutical grade; Hospira) treatment. Intraventricular injections (1.5 μl volume) of either vehicle (50% propylene glycol/50% saline) or diazepam (2 mg ml^{−1} in 50% propylene glycol) were performed in isoflurane/oxygen-anaesthetized mice (3% induction, 1–2% maintenance) using a Hamilton needle (30 gauge) and syringe (5 μl capacity) mounted to a manipulator (MP285, Sutter). The injection depth was 2.25 mm; entry site was 0.5 mm lateral from midline and immediately anterior of bregma (primary motor cortex). Two injections were made, the initial injection was made at the time of MD on the same side as the MD, and the second injection was made on the other hemisphere 24–28 h after the initial injection. Coordinates were verified before the experiment in three test animals by injecting 1,1'-diocetadecyl-3,3',3'-tetramethylindocarbocyanine percholate (DiI; 0.5 μl) and perfusing mice 1 day later. Brains were then sectioned, slices were visualized under epifluorescence and it was determined that DiI was specifically localized to ventricles using the above coordinates.

Expression of GCaMP6 and hM₄D DREADD receptors. PV-Cre mice were anaesthetized with isoflurane and core body temperature maintained as above. The scalp was retracted and a small burr hole was drilled over the left occipital pole overlying the binocular zone of primary visual cortex. GCaMP6 (ref. 24; UPenn Vector Core: AAV-1-PV2824; generously supplied by the GENIE Project, Janelia Farm Research Campus, Howard Hughes Medical Institute) and Cre-dependent hM₄D DREADD receptors^{19,20,40} (UNC Vector Core: AAV-hSyn-DIO-hM₄D(Gi)-mCherry) were expressed in cortical neurons using adeno-associated virus (AAV). Notably, hM₄D receptors were only expressed in PV-Cre cells. AAV-GCaMP6 and AAV-hM₄D were mixed at a 1:1 ratio (injected concentration 2.5×10^{11}

genome copies per ml) and loaded into a glass micropipette. The pipette was slowly inserted into the cortex and five injections were made, one each at 350, 300, 250, 200 and 150 μm below the pial surface along a single injection tract. AAV was pressure injected at each site using a PicoSpritzer: 30 puffs at 10 pounds per square inch and 10-ms duration was used at each site. Total volume injected across all five sites was roughly 0.5 μl . The scalp was then sutured shut over the burr hole and the mouse was left for 2 weeks to enable high GCaMP6 and hM₄D expression. Thereafter, a craniotomy was performed to expose a 2.5-mm-diameter region of the brain centred on binocular visual cortex, and the craniotomy was sealed with a no. 1 coverglass as described previously⁴¹.

Visual stimulation and data acquisition. To evoke responses in labelled neurons, visual stimuli were presented based on the methods of ref. 42. Neural responses to stimulation of either eye were assessed using full-field drifting square wave gratings presented at full contrast at six orientations spaced 30° apart at two directions of motion (12 orientations, temporal frequency of 2 Hz, 0.04 cycles per degree) every 11 s for a duration of 5 s preceded by a grey screen for a duration of 6 s. During the stimulation time, randomly chosen grating orientations changed every 0.4 s. Four stimulus repetitions were presented and interleaved between eye stimulation. Image sequences (256 \times 256 pixels, covering a field of view of 130 \times 130 μm) were acquired at 2.05 Hz at a depth of 180–300 μm below the pia surface using ScanImage software⁴³. Fields of view that contained no or unevenly distributed responses to ipsilateral eye stimulation, an indication that the imaged position was outside or at the border of the binocular zone, were excluded from further analysis. Neurons were distinguished from astrocytes using sulforhodamine coinjection⁴⁴. During OGB and GCaMP6 imaging, red (semrock filter: 583/22) and green (semrock filter: 510/84) emissions were separated (semrock dichroic: FF568) and detected simultaneously using two photomultiplier tubes (PMTs; Hamamatsu R3896). OGB was excited at wavelength of 870 nm. GCaMP6 was excited at 940 nm. When imaging OGB, PV interneurons were subsequently identified and excluded from analysis by switching the excitation wavelength to 960 nm. When imaging GCaMP6, PV interneurons were identified by their emission in the red channel. To facilitate defining the stimulus onset for each cell, output from a photodiode affixed to the presentation monitor was digitized simultaneously with the output from the PMTs.

For GCaMP6/hM₄D-expressing mice, images of cell responses to visual stimulation were first obtained on P35. Immediately thereafter the lid margins to the contralateral eye were sutured shut for 3 days. During the first day of lid suture mice either received two intraperitoneal injections, spaced 12 h apart of either saline or CNO (5 mg kg^{−1} dissolved in 0.9% sterile saline). On P38, the lid margins were reopened and cortical responses to visual stimulation through both the ipsilateral and contralateral eyes were assessed. Distributions of ipsilateral and contralateral response strength after MD in both conditions were compared to those obtained from baseline images.

Data analysis. A semiautomatic custom-written routine (Matlab) was used to identify the outlines of individual OGB-loaded cells³⁸. In brief, an average image was generated from the acquired time series and the x,y spatial position of a given cell was seeded by the user, blind to experimental condition. The spatial location of the cell border was then calculated based on identifying a contiguous region having a > 1.1 -fold intensity above the local average. Local average and maximum cell size was bounded by a square area $R \times R$ centred on the user defined seed pixel, where $R = 25 \mu\text{m}$. For GCaMP6 imaging, where labelled neurons have a donut appearance, we used a separate method for segmenting neurons. In brief, after manual selection of the cells to be studied, a deformable snake was used to track their individual boundaries⁴⁵, along with morphological dilation⁴⁶ to define the region of interest for each cell. For both OGB and GCaMP6 images, the pixels within each cell's region of interest were averaged to generate a single time course of fluorescent intensity for each individual cell. Response amplitude was computed from 8–12 stimulus repetitions as the first 5 frames immediately following the frame of stimulus onset divided by the average of the 6 frames preceding stimulus onset. The frame of stimulus onset was defined independently for each cell. Cells were considered responsive if the calcium transients to any of the orientation stimuli were significantly different from the baseline (ANOVA at $P < 0.01$). An ocular dominance score was calculated for all responsive cells (excluding identified PV interneurons and astrocytes) as 1 minus the ratio of the ipsilateral eye response divided by the summed response of both eyes. For each animal, a histogram of ocular dominance scores was generated and the contralateral bias index was calculated as in ref. 9, using the following formula: $((n_1 - n_7) + (2/3)(n_2 - n_6) + (1/3)(n_3 - n_5) + N)/(2N)$, in which N = total number of responsive cells and n_x is the number of cells with ocular dominance scores equal to x .

Statistical analyses and bar graphs. All data are reported as mean \pm standard error of the mean (s.e.m.). When comparing two independent groups, normally distributed data were analysed using a Student's t -test, in the case data were not normally distributed a Mann–Whitney U -test was used. In the case more than

two groups were compared and data were normally distributed, an ANOVA was performed as indicated in main text and followed by post-hoc comparisons when justified (alpha set to 0.05). Post-hoc *P* values were corrected for multiple comparisons as indicated. Bonferroni correction was used unless data had unequal variances (assessed by the Levene statistic), in that case Tamhane correction was used. In all cases, sample size *n* was defined as cell number, except in the case of comparing the contralateral bias index across treatments and GCaMP6 'optical field potential', where *n* was defined as animal number.

30. Dombeck, D. A., Khabbaz, A. N., Collman, F., Adelman, T. L. & Tank, D. W. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* **56**, 43–57 (2007).
31. Kuhlman, S. J., Tring, E. & Trachtenberg, J. T. Fast-spiking interneurons have an initial orientation bias that is lost with vision. *Nature Neurosci.* **14**, 1121–1123 (2011).
32. Xu, X., Olivas, N. D., Levi, R., Ikrar, T. & Nenadic, Z. High precision and fast functional mapping of cortical circuitry through a novel combination of voltage sensitive dye imaging and laser scanning photostimulation. *J. Neurophysiol.* **103**, 2301–2312 (2010).
33. Antonini, A., Fagiolini, M. & Stryker, M. P. Anatomical correlates of functional plasticity in mouse visual cortex. *J. Neurosci.* **19**, 4388–4406 (1999).
34. Suter, B. A. *et al.* Ephus: multipurpose data acquisition software for neuroscience experiments. *Front. Neural Circuits* **4**, 100 (2010).
35. Hooks, B. M. *et al.* Laminar analysis of excitatory local circuits in vibrissa motor and sensory cortical areas. *PLoS Biol.* **9**, e1000572 (2011).
36. Shepherd, G. M., Pologruto, T. A. & Svoboda, K. Circuit analysis of experience-dependent plasticity in the developing rat barrel cortex. *Neuron* **38**, 277–289 (2003).
37. Shi, Y., Nenadic, Z. & Xu, X. Novel use of matched filtering for synaptic event detection and extraction. *PLoS ONE* **5**, e15517 (2010).
38. Gdalyahu, A. *et al.* Associative fear learning enhances sparse network coding in primary sensory cortex. *Neuron* **75**, 121–132 (2012).
39. Stosiek, C., Garaschuk, O., Holthoff, K. & Konnerth, A. *In vivo* two-photon calcium imaging of neuronal networks. *Proc. Natl Acad. Sci. USA* **100**, 7319–7324 (2003).
40. Rogan, S. C. & Roth, B. L. Remote control of neuronal signaling. *Pharmacol. Rev.* **63**, 291–315 (2011).
41. Holtmaat, A. *et al.* Long-term, high-resolution imaging in the mouse neocortex through a chronic cranial window. *Nature Protocols* **4**, 1128–1144 (2009).
42. Mrsic-Flogel, T. D. *et al.* Homeostatic regulation of eye-specific responses in visual cortex during ocular dominance plasticity. *Neuron* **54**, 961–972 (2007).
43. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).
44. Nimmerjahn, A., Kirchhoff, F., Kerr, J. N. & Helmchen, F. Sulforhodamine 101 as a specific marker of astroglia in the neocortex *in vivo*. *Nature Methods* **1**, 31–37 (2004).
45. Xu, C. & Prince, J. L. Snakes, shapes, and gradient vector flow. *IEEE Trans. Image Process.* **7**, 359–369 (1998).
46. Shih, F. Y. *Image Processing and Mathematical Morphology: Fundamentals and Applications* (CRC, 2009).

The BC component of ABC toxins is an RHS-repeat-containing protein encapsulation device

Jason N. Busby¹, Santosh Panjikar^{2,3}, Michael J. Landsberg⁴, Mark R. H. Hurst⁵ & J. Shaun Lott¹

The ABC toxin complexes produced by certain bacteria are of interest owing to their potent insecticidal activity^{1,2} and potential role in human disease³. These complexes comprise at least three proteins (A, B and C), which must assemble to be fully toxic⁴. The carboxy-terminal region of the C protein is the main cytotoxic component⁵, and is poorly conserved between different toxin complexes. A general model of action has been proposed, in which the toxin complex binds to the cell surface via the A protein, is endocytosed, and subsequently forms a pH-triggered channel, allowing the translocation of C into the cytoplasm, where it can cause cytoskeletal disruption in both insect and mammalian cells⁵. Toxin complexes have been visualized using single-particle electron microscopy^{6,7}, but no high-resolution structures of the components are available, and the role of the B protein in the mechanism of toxicity remains unknown. Here we report the three-dimensional structure of the complex formed between the B and C proteins, determined to 2.5 Å by X-ray crystallography. These proteins assemble to form an unprecedented, large hollow structure that encapsulates and sequesters the cytotoxic, C-terminal region of the C protein like the shell of an egg. The shell is decorated on one end by a β -propeller domain, which mediates attachment of the B–C heterodimer to the A protein in the native complex. The structure reveals how C auto-proteolyses when folded in complex with B. The C protein is the first example, to our knowledge, of a structure that contains rearrangement hotspot (RHS) repeats⁸, and illustrates a marked structural architecture that is probably conserved across both this widely distributed bacterial protein family and the related eukaryotic tyrosine-aspartate (YD)-repeat-containing protein family, which includes the teneurins⁹. The structure provides the first clues about the function of these protein repeat families, and suggests a generic mechanism for protein encapsulation and delivery.

ABC toxins were first identified, and have been best characterized, in the bacterium *Photobacterium luminescens*¹. However, the entomopathogenic bacterium *Yersinia entomophaga* contains a related toxin complex locus that includes an A component encoded by two open-reading frames (ORFs) (*yenA1* and *yenA2*), a single B gene (*yenB*) and two C genes (*yenC1* and *yenC2*)¹⁰, the products of which associate independently with the A and B proteins, giving rise to two toxin complexes from one genetic locus. The C proteins of this and other ABC toxin complexes are similar to the ‘polymorphic toxins’ described previously¹¹ as they have a conserved RHS-repeat-containing amino-terminal region and a variable C-terminal region¹⁰. The C-terminal regions (CTRs) of the *Y. entomophaga* C proteins are predicted to have different toxic activities: the C1^{CTR} is homologous to cytotoxic necrotising factor 1 (CNF1) from *Escherichia coli*¹², whereas the C2^{CTR} is homologous to the deaminase YwqJ from *Bacillus subtilis*¹³. As in related complexes³, when the B subunit is co-expressed with either of the C subunits, C is cleaved at the boundary between the conserved N-terminal region and the variable C-terminal region, and both proteolysed fragments remain tightly associated with B.

To understand the role of B proteins in toxin complex activity, we prepared the complex of the *Y. entomophaga* B subunit (167 kilodaltons (kDa)) with the conserved 76-kDa N-terminal region (NTR) of the C2 subunit (C^{NTR}; see Methods for details of complex preparation) and determined its structure using X-ray crystallography (Supplementary Table 1). The structure reveals that the B and C^{NTR} proteins fold cooperatively, assembling a remarkable, intimately associated heterodimer that forms a large, hollow shell (Fig. 1a–c). An immediately notable feature is the single, long β -sheet, consisting of 76 β -strands derived from both proteins, that constitutes most of the shell structure. The shell is completed by 14 β -strands from a second β -sheet contributed by the C^{NTR}, bringing the total number of β -strands that wrap around what is a substantial central cavity to 90 (Fig. 1b, c and Supplementary Fig. 1). The C terminus of the B protein is in close proximity to the N terminus of the C protein (Supplementary Fig. 2), suggesting that the two proteins could be produced as a single polypeptide. Evidence in support of this can be found in the bacterium *Burkholderia rhizoxinica*, in which a single ORF (*tcdB2*) encodes an apparent B–C fusion protein (Supplementary Fig. 3).

The central cavity is a solvent-accessible space approximately 42 Å wide and 87 Å long, with a total enclosed volume of roughly 59,000 Å³. The shell is closed at both ends. The B end is occluded by a β -propeller domain inserted into the loop between strands β 29 and β 51. The C end is sealed by a conserved sequence defined as the ‘RHS repeat-associated core domain’ (TIGRFAM accession TIGR03696), which forms another short strip of β -sheet (strands β 45– β 49) that spirals inwards, forming a plug. The overall shape of the complex is reminiscent of a cylindrical cannister, inside of which lies the C-terminal end of C^{NTR} (that is, the cleavage site between the two portions of C). It therefore seems likely that in the complete B–C complex, the C^{CTR} is encapsulated within the shell. This proposal is supported by small-angle X-ray scattering data, which are consistent with a hollow spheroid for the B–C^{NTR} complex, but with a solid spheroid for the complete B–C complex (Supplementary Figs 4–6 and Supplementary Tables 2–5). This explains how the CTRs of C proteins remain tightly associated with the complex after auto-proteolysis, despite the absence of any covalent linkage between the proteins. In broader terms, it also explains how generally cytotoxic proteins encoded by the CTRs of C proteins, such as deaminases and proteases, can be safely expressed without causing damage to the producing cell. A plausible explanation supported by this structure would be that the toxic load remains sequestered until exposure to a change in pH triggers its release⁵ (Fig. 2a).

The amino acids immediately preceding the cleavage site in C are clearly visible in the electron density, allowing us to suggest a mechanism of auto-proteolysis. We propose that two conserved aspartate residues (Asp 663 and Asp 686), positioned either side of the last residue before cleavage (Met 690), form the catalytic site for proteolysis (Fig. 2b). In our structure, they are too far apart (6.5 Å between carboxyl oxygens) to form the canonical aspartic protease arrangement, but these residues may adopt a slightly different conformation in the pre-cleavage state. To

¹AgResearch Structural Biology Laboratory, School of Biological Sciences, The University of Auckland, Auckland 1142, New Zealand. ²Australian Synchrotron, 800 Blackburn Road, Clayton, Victoria 3168, Australia. ³Department of Biochemistry and Molecular Biology, Monash University, Victoria 3800, Australia. ⁴Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland 4072, Australia. ⁵Innovative Farming Systems, AgResearch, Lincoln Research Centre, Christchurch 8140, New Zealand.

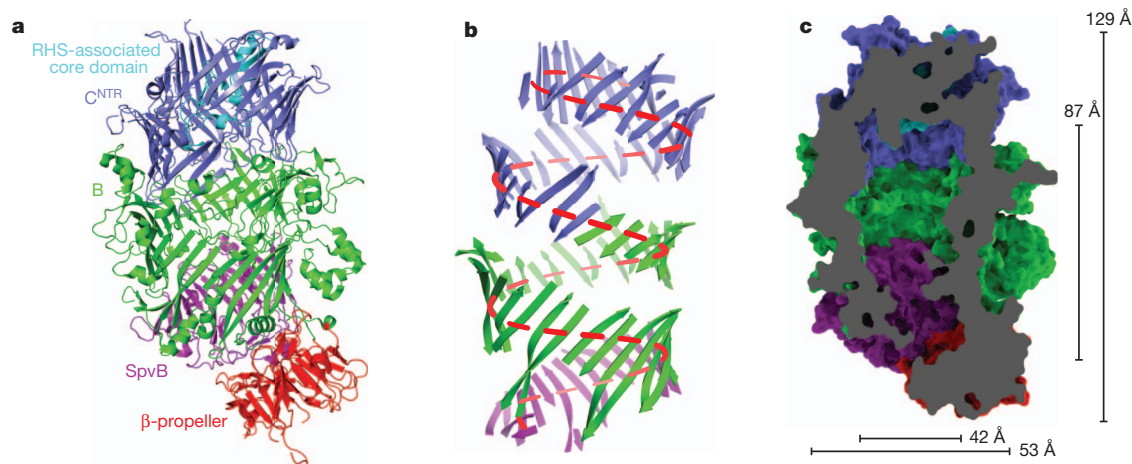


Figure 1 | Structure of the B-C^{NTR} complex. **a**, Side view of the B-C^{NTR} complex. The β -propeller domain of B is coloured red, the conserved SpvB sequence is purple, and the rest of B is green. C^{NTR} is coloured blue and the conserved RHS-associated core domain is cyan. **b**, Schematic showing how the

shell is made from a continuous strip of β -sheet, wrapping around a central cavity. **c**, A slice through the complex, with the molecular surface rendered to show the central cavity. The approximate interior and exterior dimensions are marked.

test their role, we made point mutations that replaced them with asparagines (Asp663Asn and Asp686Asn). Mutation of either residue, or of the highly conserved arginine residue (Arg 650) that is adjacent to Asp 663, completely abolished auto-proteolytic activity (Fig. 2c). These residues are highly conserved in the RHS repeat-associated core domain sequence (InterPro IPR022385; Supplementary Fig. 7) that is also present in all other known C proteins, including those from *P. luminescens*. Our structure shows that this domain, which is widely distributed across the archaea, bacteria and eukaryota¹⁴, is a cryptic aspartic protease.

RHS repeats themselves (Pfam PF05593) are present in many polymorphic toxin complexes that are found across a diverse range of bacterial species, but they are structurally uncharacterized. The structure of

the C^{NTR} represents the first example, to our knowledge, of a protein structure containing RHS repeats. Individual RHS-repeat-containing proteins can vary in size, and whereas the overall sequence conservation across the family is low, a consensus sequence for the RHS repeat itself has previously been defined: GXXXRYXYDXXGRL(I/T)¹⁵. When this is mapped onto the structure of C^{NTR} (Fig. 3 and Supplementary Fig. 8), it is clear that each RHS repeat corresponds to a single strand-turn-strand motif, many copies of which make up the extended β -sheet of the shell. Although the initial glycine is not especially well conserved, it marks the loop facing the C end of the shell. The central DXXGR motif (coloured green and blue in Fig. 3) creates the hairpin facing the B end, with the aspartic acid hydrogen-bonding to the backbone amides of the glycine and adjacent arginine. This glycine is largely conserved, but the aspartic acid can be replaced by a glutamic acid, threonine or serine residue, and typically the interactions formed remain the same. The YXY motif (coloured magenta in Fig. 3) places the two tyrosine side chains inside the shell where they sit parallel to each other. Hydrophobic amino acids at the C-terminal end of the repeat (coloured yellow in Fig. 3) also lie inside the shell on alternating β -strands. Together with the tyrosines, they form a continuous stripe along the internal face of the β -sheet that is largely hydrophobic in character.

The RHS structural motif is also present in the structure of the B subunit, albeit with less sequence conservation than that found in the C^{NTR}. The B sequence contains more insertions and extensions within

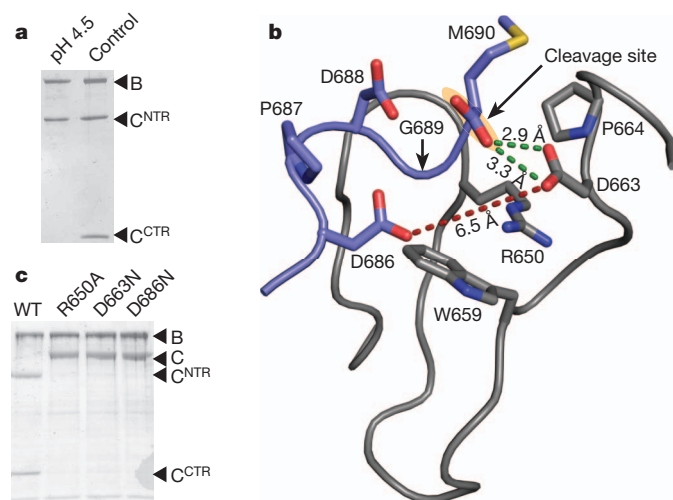


Figure 2 | Auto-proteolysis of the C protein. **a**, The effect of pH on the B-C complex. C^{CTR} dissociates from the complex and precipitates when dialysed against buffer at pH 4.5, but not at pH 7.5 (control). **b**, The site of auto-proteolysis in C. Residues immediately upstream of the cleavage point are coloured blue, surrounding residues are coloured grey. The distance between two proteolytically essential residues, Asp 686 and Asp 663, is shown in red. The side chain of Asp 663 is within hydrogen-bonding distance of the terminal carboxyl group of the cleaved peptide (distances shown in green). The amino acid immediately preceding the cleavage site, Met 690, is indicated. **c**, The effect of point mutations on C auto-proteolysis. Mutation of any one of the three highly conserved charged residues in the RHS-associated core domain (Arg 650, Asp 663 and Asp 686) eliminates proteolytic activity.

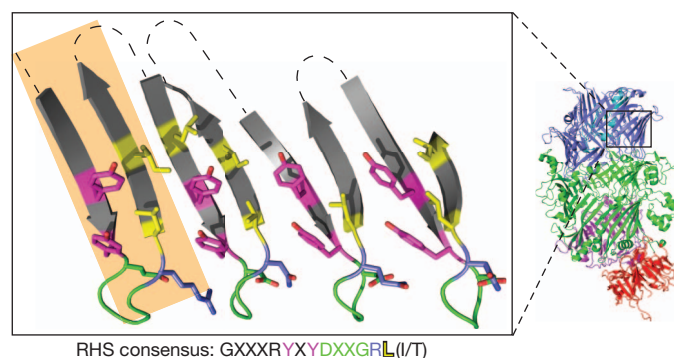


Figure 3 | RHS repeat structure. A section of the shell showing the pattern of RHS repeats, viewed from inside the central cavity. A single RHS repeat is highlighted in beige. The hairpin turn made by the DXXGR motif is coloured green and blue, and the conserved pattern of hydrophobic residues on the inside face of the β -sheet is coloured magenta (conserved tyrosines) and yellow (other hydrophobic residues).

the RHS repeats than the C sequence, which makes identifying the RHS pattern difficult by sequence conservation alone. However, inspection of the structure reveals many examples of DXXG turns, as well as tyrosine or phenylalanine side chains arranged in an equivalent fashion. Using this structural conservation as a guide, we were able to produce a refined consensus sequence for the RHS repeat (Supplementary Fig. 9a) and show that the pattern of conservation is identical to that seen in YD repeats (TIGRFAM TIGR01643; Supplementary Fig. 9b). YD repeats are found in many bacterial and eukaryotic proteins, notably in the extracellular domains of teneurin proteins, which are developmental signalling proteins conserved from flies to mammals and required for synaptic partner matching^{16,17}. We propose that RHS and YD repeats represent the same conserved structural motif that will always give rise to an extended β -sheet, forming a similar shell structure to that seen here. Support for this proposal can be found in previous low-resolution electron microscopy images of the extracellular domains of mouse teneurin, which revealed globular domains of similar dimensions to the B-C^{NTR} complex¹⁸. We predict that the YD-repeat-containing domains of eukaryotic teneurins will encapsulate their C-terminal regions, the teneurin C-terminal associated peptides (TCAPs), which are known to be active extracellular signalling components in mice^{19,20}.

Previous visualization of complete ABC toxin complexes from *Y. entomophaga*⁶ and *P. luminescens*⁷ using electron microscopy single-particle analysis has shown that the B-C complex sits in the vestibule of the channel-forming domain of A, positioned at the end of the toxin complex furthest from the membrane. The overall shape of the B-C^{NTR} dimer described here is consistent with the density seen in these studies^{6,7}, and cross-correlation of the B-C^{NTR} crystal structure with averaged electron microscopy projections of the *Y. entomophaga* toxin complex (Yen-Tc) unambiguously identified the five-bladed β -propeller domain of B as the point of interaction with the A pentamer (Fig. 4a). In Yen-Tc,

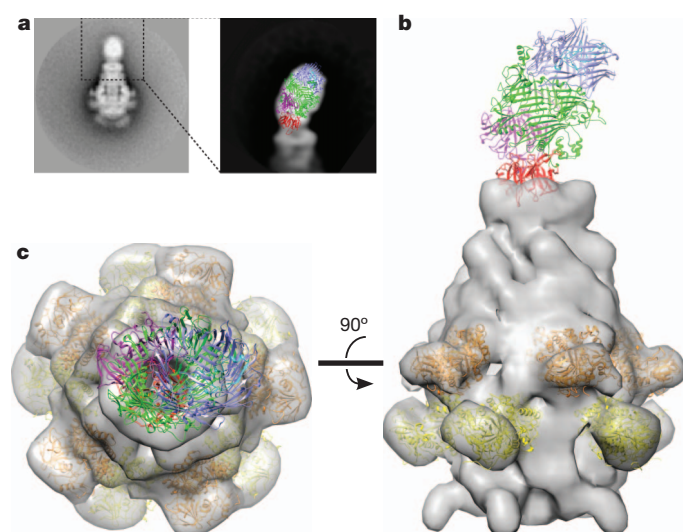


Figure 4 | Position of the B-C^{NTR} complex in the complete Yen-Tc particle. **a**, Averaged projection images obtained by negative stain electron microscopy reveal the structural conformation of the B-C complex in the context of the Yen-Tc particle. Particle projection images of Yen-Tc were aligned (left-hand box) and then masked using the region of the particle delineated by the dashed box. Class averages of the B-C complex were obtained following a further phase of classification and alignment. A representative class average is shown (right-hand box) cross-correlated with the B-C^{NTR} crystal structure, the latter shown as a ribbon diagram with the same colour scheme as Fig. 1a. **b**, **c**, Orthogonal views of the complete Yen-Tc particle. The associated chitinases Chi1 (ref. 21) and Chi2 (Protein Data Bank (PDB) accession 4DWS) are shown as orange and yellow ribbon diagrams, respectively. The electron microscopy map of the A-Chi1-Chi2 complex determined at a resolution of 17 Å by single particle analysis⁶ is shown in grey.

in which the A component is encoded by two separate ORFs, this represents an interaction with YenA2. This enables us to model a complete ABC toxin complex for the first time (Fig. 4b, c), to our knowledge, by docking the B-C^{NTR} complex and both associated chitinase enzymes²¹, Chi1 and Chi2, into the previously determined 17 Å electron microscopy structure of the *Y. entomophaga* A pentamer⁶.

A recent electron microscopy study on the *P. luminescens* ABC toxin complex⁷ has revealed the first insights into the mechanism by which toxin complexes alter their structure after interaction with lipid bilayers, allowing for injection of their toxic load through the endosomal membrane. We therefore propose a general model for the mechanism of ABC toxin complexes, on the basis of work described previously⁷ and presented here, in which the C-terminal region of C is translocated through a transmembrane pore formed by A, whereas the B and C^{NTR} proteins remain attached to the toxin complex (Supplementary Fig. 10). It remains unclear whether the cytotoxic C^{CTR} is encapsulated in a folded or unfolded state within the B-C^{NTR} shell, but the central cavity is large enough to contain it when folded (Supplementary Fig. 11). The overall architecture of the B-C^{NTR} shell, with its conserved RHS repeats producing an interior hydrophobic pattern of tyrosine, leucine and isoleucine residues, is reminiscent of the protein chaperone GroEL²². It is therefore tempting to speculate that the function of the B-C shell, and of RHS- and YD-repeat proteins more generally, may be to encapsulate unfolded proteins. There is support for this idea in the observation that many polymorphic toxins have predicted proteases as their toxic components, which would need to be contained in an inactive state to prevent proteolysis of the shell itself¹. As the pore formed by the A subunit has not yet been visualized in an active conformation, it remains unknown whether the translocating conformation of the toxin contains an open pore wide enough to allow the passage of the folded protein.

Whether it is folded or not, release of the encapsulated C^{CTR} from the B-C^{NTR} shell requires a conformational change in the latter, as there are no gaps in the structure large enough for a polypeptide to pass through. Inspection of the structure suggests that two possibilities exist: the β -propeller blades could separate, allowing extrusion of an unfolded polypeptide through the middle of the propeller, or the propeller domain could swing aside, hinged on the β 29/ β 30 and β 50/ β 51 loops that form the only covalent connections between the β -propeller and the main body of the shell. Either mechanism is likely to be dependent on both the pH-driven release mentioned previously (Fig. 2a) and mechanical interactions with the A component of the toxin. Further studies are required to establish the full details of the toxin release process.

In summary, the structure of the B-C^{NTR} complex presented here reveals how the cytotoxic C proteins of ABC-type toxin complexes are processed and protected, demonstrates the function of the B protein within the complex and provides a framework for further experiments to build a complete mechanistic model of action for this class of toxins. More broadly, it also illuminates the function of the widely distributed RHS- and YD-repeat families of proteins, which had until now been unknown.

METHODS SUMMARY

The B-C^{NTR} protein complex was obtained by co-expression of the *Y. entomophaga* B and C2 subunits in *E. coli*, followed by Ni-affinity and size-exclusion chromatography and finally dialysis of B-C2 against acetate buffer at pH 4.5. B-C^{NTR} was separated from the precipitated C2^{CTR} fragment by filtration and further size-exclusion chromatography. Crystallization of B-C^{NTR} was carried out by hanging-drop vapour diffusion with microseeding in drops containing 18% (w/v) PEG 3350, 0.15 M KH₂PO₄, pH 4.8. X-ray diffraction data were collected to a resolution of 2.5 Å at beamline MX2 at the Australian Synchrotron using the BlueIce data collection software²³, integrated using XDS²⁴ and scaled and merged using Aimless²⁵. Phasing was accomplished using SAD from a combination of Ta₂Br₁₂-soaked and selenomethionine-substituted crystals^{25–27}. Structure refinement and analysis was performed using Phenix²⁸. Electron microscopy of negatively stained Yen-Tc complexes is described elsewhere⁶. Diagrams were produced using

PyMOL (The PyMOL Molecular Graphics System, version 1.3r1, Schrödinger, LLC) and Chimera²⁹. Other procedures are described in full in the Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 25 April; accepted 16 July 2013.

Published online 4 August 2013.

1. Bowen, D. *et al.* Insecticidal toxins from the bacterium *Photobacterium luminescens*. *Science* **280**, 2129–2132 (1998).
2. French-Constant, R. H. & Waterfield, N. R. Ground control for insect pests. *Nature Biotechnol.* **24**, 660–661 (2006).
3. Pinheiro, V. B. & Ellar, D. J. Expression and insecticidal activity of *Yersinia pseudotuberculosis* and *Photobacterium luminescens* toxin complex proteins. *Cell Microbiol.* **9**, 2372–2380 (2007).
4. Waterfield, N., Hares, M., Yang, G., Dowling, A. & French-Constant, R. Potentiation and cellular phenotypes of the insecticidal toxin complexes of *Photobacterium* bacteria. *Cell Microbiol.* **7**, 373–382 (2005).
5. Lang, A. E. *et al.* *Photobacterium luminescens* toxins ADP-ribosylate actin and RhoA to force actin clustering. *Science* **327**, 1139–1142 (2010).
6. Landsberg, M. J. *et al.* 3D structure of the *Yersinia enterocolitica* toxin complex and implications for insecticidal activity. *Proc. Natl Acad. Sci. USA* **108**, 20544–20549 (2011).
7. Gatsogiannis, C. *et al.* A syringe-like injection mechanism in *Photobacterium luminescens* toxins. *Nature* **495**, 520–523 (2013).
8. Hill, C. W., Sandt, C. H. & Vlazny, D. A. Rhs elements of *Escherichia coli*: a family of genetic composites each encoding a large mosaic protein. *Mol. Microbiol.* **12**, 865–871 (1994).
9. Minet, A. D., Rubin, B. P., Tucker, R. P., Baumgartner, S. & Chiquet-Ehrismann, R. Teneurin-1, a vertebrate homologue of the *Drosophila* pair-rule gene ten-m, is a neuronal protein with a novel type of heparin-binding domain. *J. Cell Sci.* **112**, 2019–2032 (1999).
10. Hurst, M. R. H. *et al.* The main virulence determinant of *Yersinia enterocolitica* MH96 is a broad-host-range toxin complex active against insects. *J. Bacteriol.* **193**, 1966–1980 (2011).
11. Zhang, D., de Souza, R. F., Anantharaman, V., Iyer, L. M. & Aravind, L. Polymorphic toxin systems: comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol. Direct* **7**, 18 (2012).
12. Buetow, L., Flatau, G., Chiu, K., Boquet, P. & Ghosh, P. Structure of the Rho-activating domain of *Escherichia coli* cytotoxic necrotizing factor 1. *Nature Struct. Biol.* **8**, 584–588 (2001).
13. Iyer, L. M., Zhang, D., Rogozin, I. B. & Aravind, L. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res.* **39**, 9473–9497 (2011).
14. Jackson, A. P., Thomas, G. H., Parkhill, J. & Thomson, N. R. Evolutionary diversification of an ancient gene family (*rhs*) through C-terminal displacement. *BMC Genomics* **10**, 584 (2009).
15. Wang, Y. D., Zhao, S. & Hill, C. W. *Rhs* elements comprise three subfamilies which diverged prior to acquisition by *Escherichia coli*. *J. Bacteriol.* **180**, 4102–4110 (1998).
16. Mosca, T. J., Hong, W., Dani, V. S., Favaloro, V. & Luo, L. Trans-synaptic Teneurin signalling in neuromuscular synapse organization and target choice. *Nature* **484**, 237–241 (2012).
17. Hong, W., Mosca, T. J. & Luo, L. Teneurins instruct synaptic partner matching in an olfactory map. *Nature* **484**, 201–207 (2012).
18. Feng, K. *et al.* All four members of the Ten-m/Odz family of transmembrane proteins form dimers. *J. Biol. Chem.* **277**, 26128–26135 (2002).
19. Chand, D. *et al.* C-terminal processing of the teneurin proteins: Independent actions of a teneurin C-terminal associated peptide in hippocampal cells. *Mol. Cell Neurosci.* **52**, 38–50 (2013).
20. Tucker, R. P. & Chiquet-Ehrismann, R. Teneurins: a conserved family of transmembrane proteins involved in intercellular signaling during development. *Dev. Biol.* **290**, 237–245 (2006).
21. Busby, J. N. *et al.* Structural analysis of Chi1 chitinase from *Yen-Tc*: the multisubunit insecticidal ABC toxin complex of *Yersinia enterocolitica*. *J. Mol. Biol.* **415**, 359–371 (2012).
22. Xu, Z., Horwich, A. L. & Sigler, P. B. The crystal structure of the asymmetric GroEL–GroES–(ADP)₇ chaperonin complex. *Nature* **388**, 741–750 (1997).
23. McPhillips, T. M. *et al.* Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron Radiat.* **9**, 401–406 (2002).
24. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
25. The CCP4 suite. Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
26. Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. Auto-rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment. *Acta Crystallogr. D* **61**, 449–457 (2005).
27. Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr. D* **65**, 1089–1097 (2009).
28. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
29. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to M. Sullivan, R. Kingston, T. Baker and many other members of the Structural Biology section at the University of Auckland for discussions, to V. Arcus for initial enthusiasm, and to B. Hankamer for provision of laboratory space at the University of Queensland and for mentorship. This work was supported by the New Zealand Foundation for Research, Science and Technology contract C10X0804, awarded to M.R.H.H. We would like to thank all beamline staff at the MX1, MX2 and SAXS/WAXS beamlines at the Australian Synchrotron for their support, and to the New Zealand Synchrotron Group Ltd for synchrotron access arrangements.

Author Contributions J.N.B. cloned constructs, expressed, purified and crystallized proteins, collected and processed X-ray crystallography and SAXS data, and refined and analysed the protein structure; S.P. determined the protein structure; M.J.L. processed, refined and analysed negative-stain electron microscopy data; M.R.H.H. and J.S.L. designed the study; J.N.B. and J.S.L. wrote the paper and all authors discussed the results and contributed to the manuscript.

Author Information The atomic coordinates of the B–C^{NTR} complex and the Chi2 chitinase have been deposited in the Protein Data Bank (<http://www.pdb.org>) under the accession codes 4IGL and 4DWS, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.S.L. (s.lott@auckland.ac.nz) or M.R.H.H. (mark.hurst@agresearch.co.nz).

METHODS

C^{CTR} dissociates at low pH. When B and either C1 or C2 were co-expressed in *E. coli*, C1 and C2 auto-proteolysed into two fragments. In both cases, although all three protein fragments co-eluted as a single complex when purified by size-exclusion chromatography (SEC), we were unable to crystallize the purified complexes. As the current model for toxin complex cell entry involves exposure to low pH in the acidified endosome⁵, we tested the behaviour of the B–C1 and B–C2 complexes under a range of pH conditions from 4.5 to 9.5. At low pH (4.5–5.0), the complexes began to precipitate, with the C-terminal regions of the C proteins showing differential precipitation, allowing purification of complexes containing just B and the N-terminal portions of C. The complex used in the subsequent experiments described in this paper was prepared by co-expression of B and C2 initially, but for simplicity is referred to as B–C^{NTR} throughout, as the N-terminal regions of C1 and C2 share a high degree of sequence conservation and their overall structures are expected to be similarly well conserved.

Purification and crystallization. Selenomethionine-substituted protein was produced by expression of B–C2 in lysogeny broth (LB) medium supplemented with selenomethionine, and was purified in a similar way to the native protein. The complex was purified by immobilized metal affinity chromatography (IMAC), polyhistidine tag cleavage, subtractive IMAC and SEC. The C2^{CTR} was removed by dialysis against acetate buffer at pH 4.5, filtration and SEC. The remaining B–C^{NTR} complex was then dialysed against the original buffer at pH 7.5. Crystallization was carried out by hanging-drop vapour diffusion with microseeding in drops containing 18% (w/v) PEG 3,350, 0.15 M KH₂PO₄, pH 4.8. Ta₆Br₁₂ derivatization was carried out by transferring crystals of B–C2^{NTR} into a solution of 18% (w/v) PEG 3,350, 0.15 M KH₂PO₄, 5% (v/v) glycerol, 2 mM Ta₆Br₁₂ and incubating for 3 h, by which time the crystals had taken on a green colouration. Crystals were cryoprotected by soaking in mother liquor containing 20% glycerol and plunge-cooled in liquid nitrogen.

X-ray diffraction data collection, structure solution and refinement. Diffraction data were collected at 100 K at the MX1 (bending magnet) and MX2 (microfocus) beamlines of the Australian Synchrotron using the BlueIce data collection software²³. Data were integrated using XDS²⁴ and scaled and merged using Aimless²⁵ (Supplementary Table 1). X-ray diffraction data for native crystals were collected to a resolution of 2.5 Å at the MX2 beamline at a wavelength of 0.9537 Å. To enhance the anomalous signal, multiple data sets were collected at a wavelength of 0.9791 Å (above the absorption edge of selenium) at the MX1 beamline from the selenomethionine-substituted crystals. Wedges of 68–360° were collected from four crystals with a resolution between 2.8 and 2.9 Å. Despite the higher redundancy of the data set collected from selenomethionine-substituted crystals, the selenium substructure could not be solved. Most likely, these crystals suffered from radiation damage. Therefore, a single data set from a Ta₆Br₁₂-soaked crystal was collected at a wavelength of 1.258 Å (above the tantalum absorption edge) to 3.15 Å resolution at the MX1 beamline. The substructure of the tantalum bromide cluster could be solved easily, but the SAD phases were poor owing to the limited resolution of the derivative. The Ta₆Br₁₂-soaked crystal turned out to be non-isomorphous to the selenomethionine-substituted crystals. The poor SAD phases from the Ta₆Br₁₂-soaked crystal were transferred to the selenomethionine-substituted crystals by the multiple crystal averaging method. The resulting phase

information was used to determine a substructure of the selenomethionine derivative by the anomalous difference Fourier method. The crystal structure was then solved using selenium-SAD phasing^{25–27}. The resulting structure was used to phase the higher-resolution native data by molecular replacement using Phenix²⁸, resulting in high-quality electron density maps that were readily interpretable (Supplementary Fig. 12). Structure refinement and analysis was performed using Phenix²⁸ and diagrams were produced using PyMOL (The PyMOL Molecular Graphics System, version 1.3rl, Schrödinger, LLC) and Chimera²⁹. The final structure had 96% of backbone dihedral angles lying within the favoured region of the Ramachandran plot and 0.12% outliers.

SAXS data collection and processing. Small-angle X-ray scattering (SAXS) data were collected at the SAXS/WAXS beamline at the Australian Synchrotron (Supplementary Figs 4–6 and Supplementary Tables 2–5). Samples were purified to homogeneity by IMAC and SEC, and exhaustively dialysed against sample buffer containing 20 mM HEPES, pH 7.5, 150 mM NaCl. The dialysate was used as the solvent blank. Data collection was carried out at 291 K with 1 or 2 s exposures. Sample was flowed across the beam at 4 µl s^{−1} to avoid radiation damage. Multiple concentrations were tested for each protein, and images for each concentration were compared, averaged and buffer-subtracted using the ScatterBrain software provided by the Australian Synchrotron. Scattering data were placed on the absolute scale by measuring the scattering of a water sample³⁰. *Ab initio* bead models were created by running dammif³¹ 20 times, superimposing and averaging the resulting models with damaver³², and using this as input for a final refinement run of dammin³³. SAXS data were compared with the theoretical scattering of the B–C^{NTR} crystal structure using crysol³⁴.

Electron microscopy and image processing. Yen-Tc was previously analysed by negative-stain transmission electron microscopy⁶. Reference-free class averages, generated from a data set of 10,224 particle projection images had initially correlated the B–C complex with a diffuse, spheroidal density attached at one end of the particle. To resolve this structure in more detail, it was necessary to remove the influence of the pentameric A particle on the alignment and classification process. This was achieved by first aligning then masking out the regions of the raw particle images corresponding to the B–C subcomplex. The masked B–C ‘subparticles’ were then iteratively re-classified and aligned using maximum likelihood analysis as implemented in Xmipp³⁵.

30. Orthon, D., Bergmann, A. & Glatter, O. SAXS experiments on absolute scale with Kratky systems using water as a secondary standard. *J. Appl. Crystallogr.* **33**, 218–225 (2000).
31. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).
32. Volkov, V. V. & Svergun, D. I. Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864 (2003).
33. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879–2886 (1999).
34. Svergun, D., Barberato, C. & Koch, M. H. J. CRY SOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).
35. Scheres, S. H. W., Nuñez-Ramírez, R., Sorzano, C. O. S., Carazo, J. M. & Marabini, R. Image processing for electron microscopy single-particle analysis using Xmipp. *Nature Protocols* **3**, 977–990 (2008).

Characterization of H7N9 influenza A viruses isolated from humans

Tokiko Watanabe^{1*}, Maki Kiso^{2*}, Satoshi Fukuyama^{1*}, Noriko Nakajima^{3*}, Masaki Imai^{4*}, Shinya Yamada², Shin Murakami⁵, Seiya Yamayoshi², Kiyoko Iwatsuki-Horimoto², Yoshihiro Sakoda⁶, Emi Takashita⁴, Ryan McBride⁷, Takeshi Noda², Masato Hatta⁸, Hirotaka Imai⁸, Dongming Zhao¹, Noriko Kishida⁴, Masayuki Shirakura⁴, Robert P. de Vries⁷, Shintaro Shichinohe⁶, Masatoshi Okamatsu⁶, Tomokazu Tamura⁶, Yuriko Tomita¹, Naomi Fujimoto¹, Kazue Goto¹, Hiroaki Katsura², Eiryo Kawakami¹, Izumi Ishikawa¹, Shinji Watanabe^{1,9}, Mutsumi Ito², Yuko Sakai-Tagawa², Yukihiko Sugita², Ryuta Uraki², Reina Yamaji², Amie J. Einfeld⁸, Gongxun Zhong⁸, Shufang Fan⁸, Jihui Ping⁸, Eileen A. Maher⁸, Anthony Hanson⁸, Yuko Uchida¹⁰, Takehiko Saito¹⁰, Makoto Ozawa^{11,12}, Gabriele Neumann⁸, Hiroshi Kida^{6,13}, Takato Odagiri⁴, James C. Paulson⁷, Hideki Hasegawa³, Masato Tashiro⁴ & Yoshihiro Kawaoka^{1,2,5,8,14}

Avian influenza A viruses rarely infect humans; however, when human infection and subsequent human-to-human transmission occurs, worldwide outbreaks (pandemics) can result. The recent sporadic infections of humans in China with a previously unrecognized avian influenza A virus of the H7N9 subtype (A(H7N9)) have caused concern owing to the appreciable case fatality rate associated with these infections (more than 25%), potential instances of human-to-human transmission¹, and the lack of pre-existing immunity among humans to viruses of this subtype. Here we characterize two early human A(H7N9) isolates, A/Anhui/1/2013 (H7N9) and A/Shanghai/1/2013 (H7N9); hereafter referred to as Anhui/1 and Shanghai/1, respectively. In mice, Anhui/1 and Shanghai/1 were more pathogenic than a control avian H7N9 virus (A/duck/Guinea/466/2011 (H7N9); Dk/GM466) and a representative pandemic 2009 H1N1 virus (A/California/4/2009 (H1N1pdm09); CA04). Anhui/1, Shanghai/1 and Dk/GM466 replicated well in the nasal turbinates of ferrets. In nonhuman primates, Anhui/1 and Dk/GM466 replicated efficiently in the upper and lower respiratory tracts, whereas the replicative ability of conventional human influenza viruses is typically restricted to the upper respiratory tract of infected primates. By contrast, Anhui/1 did not replicate well in miniature pigs after intranasal inoculation. Critically, Anhui/1 transmitted through respiratory droplets in one of three pairs of ferrets. Glycan arrays showed that Anhui/1, Shanghai/1 and A/Hangzhou/1/2013 (H7N9) (a third human A(H7N9) virus tested in this assay) bind to human virus-type receptors, a property that may be critical for virus transmissibility in ferrets. Anhui/1 was found to be less sensitive in mice to neuraminidase inhibitors than a pandemic H1N1 2009 virus, although both viruses were equally susceptible to an experimental antiviral polymerase inhibitor. The robust replicative ability in mice, ferrets and nonhuman primates and the limited transmissibility in ferrets of Anhui/1 suggest that A(H7N9) viruses have pandemic potential.

Influenza A virus infections place a considerable burden on public health and the world economy. In March 2013, several individuals were reported to be infected with an avian A(H7N9) virus^{1,2}. Viruses of this subtype do not circulate in humans, so A(H7N9) viruses capable

of transmitting among humans would encounter populations that lack any protective immunity to them. By May 30 2013, 132 confirmed human infections with A(H7N9) viruses had been reported, with 37 deaths³, resulting in a case fatality rate of >25%.

Sequence and phylogenetic analysis revealed that the haemagglutinin (HA) and neuraminidase (NA) genes of the A(H7N9) viruses originated from avian H7 and N9 viruses, respectively^{2,4,5}, whereas the remaining six genes are closely related to H9N2 subtype viruses that have circulated in poultry in China^{2,4,5}. Several of the A(H7N9) viruses possess amino acid changes known to facilitate infection of mammals, such as leucine at position 226 of HA (H3 HA numbering), which confers increased binding to human-type receptors⁶, and the mammalian-adapting mutations E627K^{7,8} or D701N⁹ in the PB2 polymerase subunit. Notably, the PB2-627K or PB2-701N markers have been detected in almost all human, but not avian or environmental, A(H7N9) isolates, suggesting ready adaptation of A(H7N9) viruses to humans.

To characterize the biological properties and pandemic potential of A(H7N9) viruses, we compared Anhui/1 (which possesses the mammalian-adapting HA-226L and PB2-627K markers) and Shanghai/1 (which possesses the 'avian-type' HA-226Q and mammalian-adapting PB2-627K markers) with the phylogenetically unrelated avian H7N9 Dk/GM466 virus, and with CA04, an early, representative 2009 H1N1 pandemic virus. Anhui/1, Shanghai/1 and CA04 replicated efficiently in Madin-Darby canine kidney (MDCK) cells and in differentiated normal human bronchial epithelial cells compared with Dk/GM466, especially at 33 °C, a temperature corresponding to the human upper airway (Supplementary Fig. 1). Electron microscopic analysis showed Anhui/1 as a spherical particle that appeared to be efficiently released from infected cells (Supplementary Fig. 2).

Next, we assessed the pathogenicity of Anhui/1 and Shanghai/1 in established animal models in influenza virus research, namely mice, ferrets and nonhuman primates (Anhui/1 only). In BALB/c mice, Anhui/1 and Shanghai/1 were more pathogenic than CA04 and Dk/GM466 on the basis of MLD₅₀ (mouse lethal dose 50; the dose required to kill 50% of infected mice) values, which were 10^{3.5} plaque-forming units (p.f.u.) for Anhui/1 and Shanghai/1, 10^{5.5} p.f.u. for CA04 and 10^{6.7} p.f.u. for Dk/GM466 (Supplementary Fig. 3). Three days post-infection

¹ERATO Infection-Induced Host Responses Project, Japan Science and Technology Agency, Saitama 332-0012, Japan. ²Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. ³Department of Pathology, National Institute of Infectious Diseases, Shinjuku, Tokyo 162-8640, Japan. ⁴Influenza Virus Research Center, National Institute of Infectious Diseases, Musashimurayama, Tokyo 208-0011, Japan. ⁵Department of Special Pathogens, International Research Center for Infectious Diseases, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. ⁶Laboratory of Microbiology, Department of Disease Control, Graduate School of Veterinary Medicine, Hokkaido University, Sapporo 060-0818, Japan. ⁷The Scripps Research Institute, 10550 North Torrey Pines Road, SP-3 La Jolla, California 92037, USA. ⁸Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin-Madison, 575 Science Drive, Madison, Wisconsin 53711, USA. ⁹Laboratory of Veterinary Microbiology, Department of Veterinary Sciences, University of Miyazaki, Miyazaki 889-2192, Japan. ¹⁰Influenza and Prion Disease Research Center, National Institute of Animal Health, Kannondai 3-1-5, Tsukuba, Ibaraki 305-0856, Japan. ¹¹Laboratory of Animal Hygiene, Joint Faculty of Veterinary Medicine, Kagoshima University, Kagoshima 890-0065, Japan. ¹²Transboundary Animal Disease Center, Joint Faculty of Veterinary Medicine, Kagoshima University, Kagoshima 890-0065, Japan. ¹³Research Center for Zoonosis Control, Hokkaido University, Sapporo 001-0020, Japan. ¹⁴Laboratory of Bioresponses Regulation, Department of Biological Responses, Institute for Virus Research, Kyoto University, Kyoto 606-8507, Japan.

*These authors contributed equally to this work.

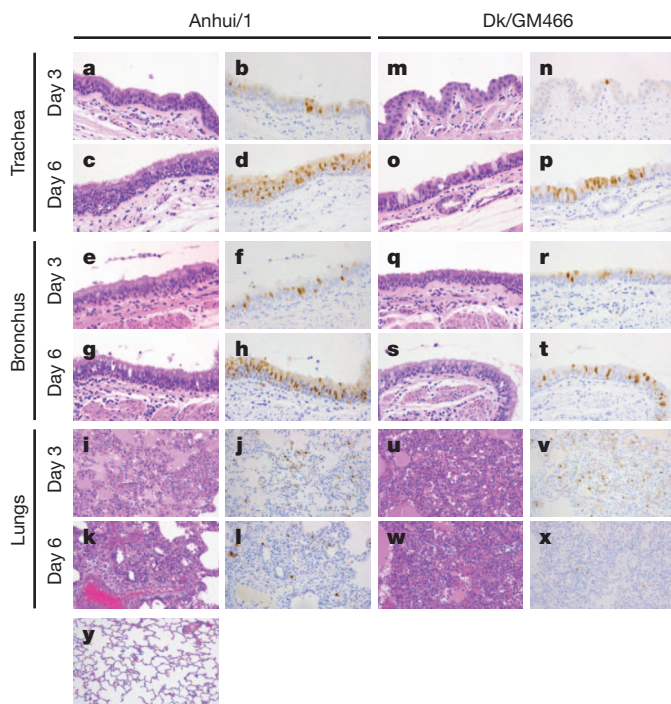


Figure 1 | Pathological findings in infected macaques. a–x, Shown are pathological findings in the trachea (a–d, m–p), bronchus (e–h, q–t) and lungs (i–l, u–x) of macaques infected with Anhui/1 (a–l, u–x) or Dk/GM466 (m–x) at 3 d.p.i. (a, b, e, f, i, j, m, n, q, r, u, v) or 6 d.p.i. (c, d, g, h, k, l, o, p, s, t, w, x), with haematoxylin and eosin staining (a, c, e, g, i, k, m, o, q, s, u, w) or immunohistochemistry for influenza viral antigen (b, d, f, h, j, l, n, p, r, t, v, x). y, Haematoxylin and eosin staining of the lung of an uninfected macaque is shown. Original magnifications: ×400 (a–h, m–t), ×200 (i–l, u–y).

(d.p.i.), virus titres in the lungs and nasal turbinates of Anhui/1-, Shanghai/1- and CA04-infected mice were slightly higher than those in Dk/GM466-infected mice (Supplementary Table 1). Lung lesions in Anhui/1- and CA04-infected mice were more severe than those in Dk/GM466-infected mice, in particular at 6 d.p.i. (Supplementary Fig. 4). Bronchitis, bronchiolitis, thickening of the alveolar septa, oedema and interstitial inflammatory cell infiltration were also more prominent in Anhui/1- and CA04-infected mice. Viral antigen was detected in many alveolar and bronchial epithelial cells at 3 d.p.i. in Anhui/1- and CA04-infected mice (Supplementary Fig. 4), whereas viral-antigen-positive cells were restricted to a few bronchial epithelial cells in Dk/GM466-infected mice (Supplementary Fig. 4). Collectively, these findings demonstrate that Anhui/1 is as pathogenic as CA04 and more pathogenic than Dk/GM466 in mice.

Ferrets intranasally infected with Anhui/1, Shanghai/1, CA04 or Dk/GM466 experienced loss of appetite. Transient weight loss was detected in one of the three animals infected with Anhui/1 (Supplementary Fig. 5). Virus titres in the trachea of Anhui/1-, Shanghai/1- and CA04-infected ferrets were higher at 3 d.p.i. than those obtained from Dk/GM466-infected animals (Supplementary Table 2); at 6 d.p.i., virus was isolated from the trachea of Anhui/1-, Shanghai/1- and Dk/GM466-infected animals, but not from that of CA04-infected ferrets. All three viruses replicated inefficiently in the lungs of these animals. In this study, we did not detect virus in the lungs in the CA04-infected animals at 3 d.p.i. (Supplementary Table 2), whereas this virus was recovered from two out of three ferrets infected at 3 d.p.i. in our previous study¹⁰, consistent with efficient replication of pandemic 2009 H1N1 virus in the lungs of ferrets as reported by others^{11–14}. Appreciable amounts of virus were recovered from nasal turbinates at 3 and 6 d.p.i., with the exception of the nasal turbinates of CA04-infected animals at 6 d.p.i. (Supplementary Table 2). Inflammation was prominent in the trachea and submucosal glands of CA04-infected ferrets (Supplementary Fig. 6). Viral-antigen-positive cells were detected in the tracheal, glandular and alveolar epithelial cells of all ferrets.

Anhui/1- and CA04-infected ferrets displayed numerous viral-antigen-positive cells, especially in the glandular epithelia, whereas Dk/GM466-infected ferrets presented far fewer antigen-positive cells. Viral antigens were also detected in pneumocytes in localized lung lesions of each ferret. In mediastinal lymph nodes, viral antigen was detected in Anhui/1- and CA04-infected ferrets only. Anhui/1 and Shanghai/1 thus established a robust, although relatively mild, infection in the upper respiratory organs of ferrets that was unlike most avian H5N1 influenza virus infections in ferrets, which consistently cause severe symptoms including profound weight loss.

Infection of cynomolgus macaques (*Macaca fascicularis*) with 6.7×10^7 p.f.u. of Anhui/1 or Dk/GM466 caused fever (Supplementary Fig. 7), as did infection with CA04 in our previous study¹⁰. Both Anhui/1 and Dk/GM466 replicated appreciably in macaque nasal turbinates, trachea and lungs, although variability among the virus titres was noticed, as is commonly found among outbred animals (Supplementary Tables 3 and 4a). Previously, we detected efficient replication of CA04 in the respiratory organs of cynomolgus macaques at 3 d.p.i.; by 7 d.p.i., no virus was detected in the lungs¹⁰.

Pathological examination of the respiratory organs did not reveal major differences between macaques infected with Anhui/1 or Dk/GM466; no notable lesions were detected in the trachea or lobular bronchus, but alveolar spaces contained oedematous exudate and inflammatory infiltrates comprising mainly neutrophils and monocytes/macrophages (Fig. 1). At 6 d.p.i., regenerative changes were observed. Lung inflammation scores did not reveal appreciable differences between animals infected with Anhui/1 or Dk/GM466 (Supplementary Table 4b). Numerous tracheal and bronchial epithelial cells of Anhui/1-infected macaques were positive for viral antigen (Fig. 1b, d, f, h), as was observed in our previous study with CA04 at 7 d.p.i.¹⁰. Fewer antigen-positive cells in the tracheal and bronchial epithelia of Dk/GM466-infected animals were detected, in particular

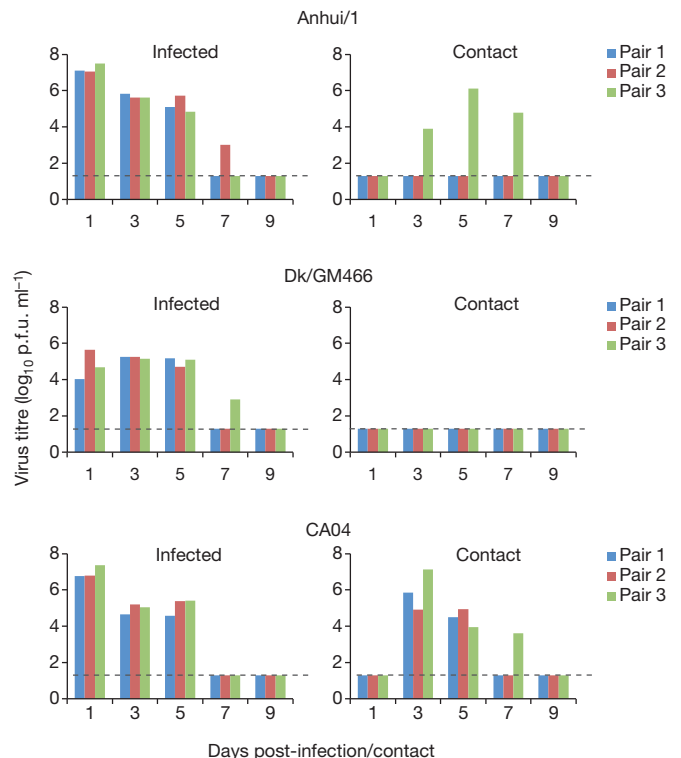


Figure 2 | Respiratory droplet transmission in ferrets. Ferrets were infected with 5×10^5 p.f.u. of Anhui/1, Dk/GM466 or CA04 (infected ferrets). One day later, three naive ferrets (contact ferrets) were each placed in a cage adjacent to an infected ferret. Nasal washes were collected from infected ferrets on day 1 after inoculation and from contact ferrets on day 1 after co-housing, and then every other day (up to 9 days) for virus titration.

at 3 d.p.i. (Fig. 1n, r). Viral antigen was also detected in the mediastinal lymph node sections of Anhui/1-infected monkeys (Supplementary Fig. 8). In addition, our analysis of chemokine/cytokine responses suggests that Anhui/1 induces strong inflammatory responses both systemically and at the site of virus infection (Supplementary Fig. 9 and Supplementary Information).

No sustained human-to-human transmission of the novel A(H7N9) viruses has been reported to date. To assess the transmissibility of Anhui/1, naive 'contact' ferrets in wireframe cages (which prevent direct contact with animals in neighbouring cages, but allow respiratory droplet transmission) were placed adjacent to an 'infected' ferret the day after infection as described previously¹⁵. We recovered viruses from the nasal washes of all three contact ferrets for CA04 (Fig. 2), as expected on the basis of previous studies by us¹⁰ and others^{13,14}. No virus was detected in the nasal washes of contact ferrets for Dk/GM466 (Fig. 2), consistent with the general lack of transmissibility of avian influenza viruses in ferrets. However, one of three contact ferrets for Anhui/1 shed virus on days 3–7 after contact (Fig. 2). Serum antibody titres against Anhui/1 confirmed infection of this animal, whereas the other two contact animals did not seroconvert (Supplementary Table 5). Given that avian H5N1 influenza viruses require several mutations to transmit through respiratory droplets among ferrets^{15–18}, the pandemic potential of A(H7N9) viruses may be greater than that of the highly pathogenic avian H5N1 influenza viruses.

Because replication and transmission of an H5 HA-possessing virus in ferrets is associated with amino acid changes in HA^{15,18}, we sequenced the genomes of viruses obtained from infected and contact ferrets. Compared to virus inoculum, we detected three non-synonymous mutations in HA (T71I, R131K and A135T; Fig. 3 and Supplementary Table 6) and one nonsynonymous mutation in NA (A27T; N9 numbering) (Supplementary Table 6); we also detected several synonymous nucleotide changes. The amino acid mutations in HA were detected in >40% of the molecular clones derived from samples obtained on day 3 post-contact (Supplementary Table 7), but were found exclusively by day 5 post-contact. Interestingly, the egg-grown virus stock of Anhui/1 possessed a mixture of amino acids at five positions, and the three amino acid changes in HA detected in the transmitted virus were detected in two of 39 HA clones representing the egg-grown virus stock of Anhui/1 (Supplementary Table 7). The selection of these mutations during A/Anhui replication and/or transmission in ferrets strongly suggests that these mutations have a biological role, possibly in HA stability and/or receptor-binding specificity or affinity. In fact, positions 131 and 135 are located near the receptor-binding pocket (Fig. 3). At position 71, both threonine and isoleucine are commonly found among H7 HAs, and the location of this position 'underneath' the receptor-binding pocket suggests a possible effect on HA stability (Fig. 3).

Our data indicate that Anhui/1 efficiently infects mammalian cells. We speculated that amino acid changes in Anhui/1 HA contribute to this host tropism. The HA-226L residue found in Anhui/1 is known to increase the affinity of H3 HAs (which are phylogenetically closely related to H7 HAs) to sialic acids linked to galactose by an α 2,6-linkage ('human-type' receptors)⁶; by contrast, avian influenza viruses preferentially bind to sialic acids linked to galactose by an α 2,3-linkage ('avian-type' receptors). In addition, Anhui/1 HA deviates from the avian virus consensus sequence at positions 186 and 189, which influence the receptor-binding preferences of H5 and H9 HAs, respectively^{19–21}.

To analyse receptor-binding preference, we subjected recombinant viruses possessing the Anhui/1, Shanghai/1 or Hangzhou/1 HA genes (see Supplementary Table 8) in combination with Anhui/1 NA genes and the remaining genes from A/Puerto Rico/8/34 (H1N1) (a laboratory-adapted strain) to glycan array analysis. All three viruses bound to α 2,6-linked sialosides, unlike a representative avian virus (A/Vietnam/1203/2004 (H5N1); Vietnam/1203), which showed typical specificity for α 2,3-linked sialosides (Fig. 3 and Supplementary Table 9). Anhui/1 and Hangzhou/1 HAs bound most strongly to α 2,6-linked sialosides, and in particular to extended N-linked glycans that are found on human

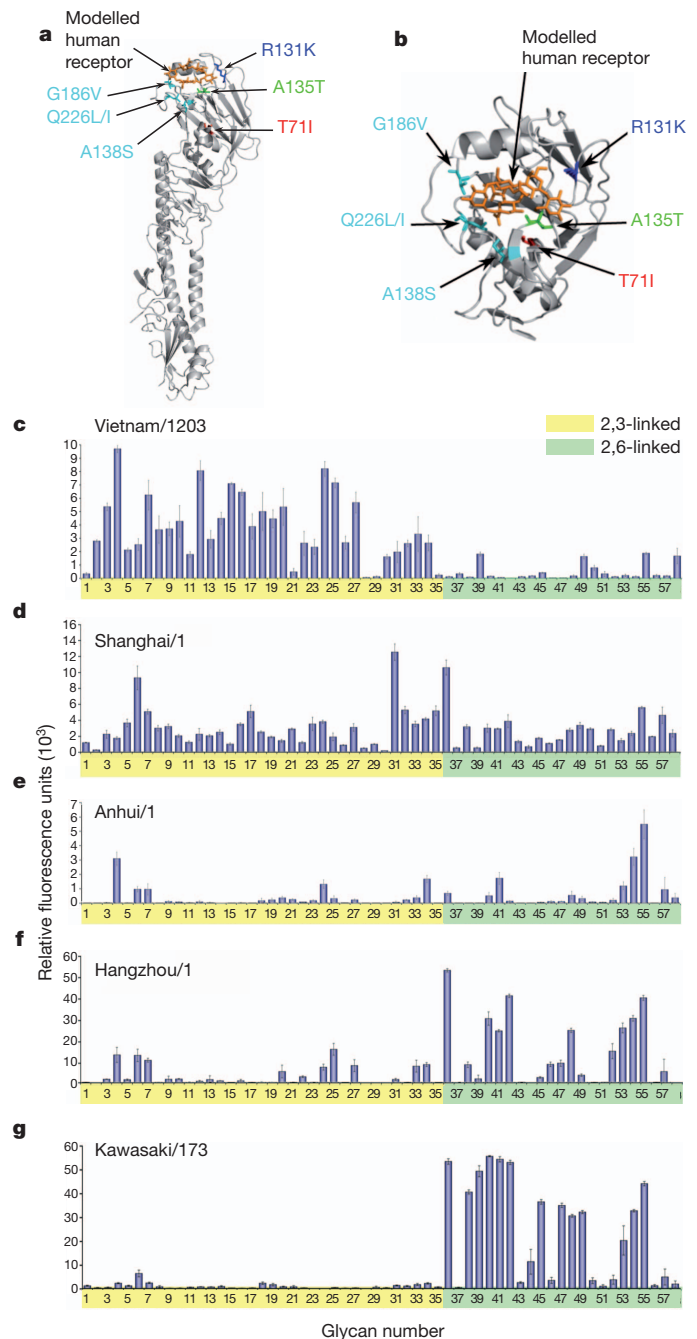


Figure 3 | HA structural analysis and glycan microarray analysis.

a, Localization of amino acid changes in a virus from a ferret infected via respiratory droplets. Shown is the three-dimensional structure of A/Netherlands/219/2003 (H7N7) HA (Protein data bank (PDB) code, 4DJ6) in complex with human receptor analogues. **b**, Close-up view of the globular head. Mutations that increase affinity to human-type receptors are shown in cyan. Mutations that emerged in Anhui/1 HA during replication and/or transmission in ferrets are shown in red (T71I), green (A135T) and blue (R131K). The human receptor analogue (derived from its complex with H9 HA (PDB, 1JSI); shown in orange) is docked into the structure. Images were created with MacPymol (<http://www.pymol.org/>). **c–g**, Receptor specificities of recombinant viruses possessing A(H7N9) HAs (Anhui/1, Shanghai/1, Hangzhou/1) were compared with representative avian (A/Vietnam/1203/2004 (H5N1); Vietnam/1203) and human (A/Kawasaki/173/2001 (H1N1); Kawasaki/173) isolates in a glycan microarray containing α 2,3 and α 2,6 sialosides. Error bars represent standard deviations calculated from six replicate spots of each glycan. A complete list of glycans is found in Supplementary Table 9.

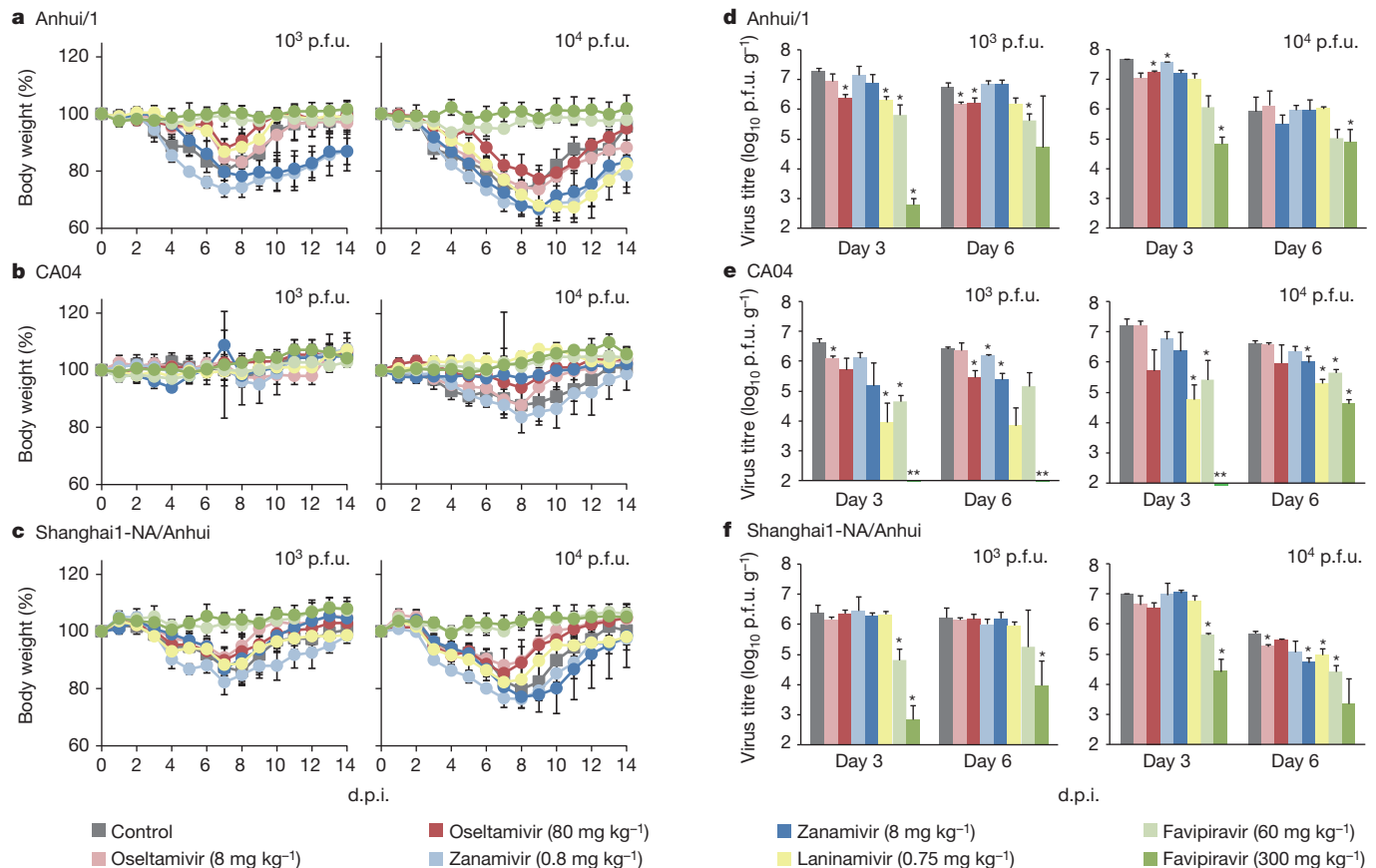


Figure 4 | Virus sensitivity to antivirals in mice. a–f, Mice were intranasally inoculated with 10^3 or 10^4 p.f.u. (50 μ l) of Anhui/1 (a, d), CA04 (b, e) or recombinant Anhui/1 virus possessing Shanghai/1-NA-294K (c, f). At 2 h after infection, mice were treated with oseltamivir phosphate, zanamivir, laninamivir octanoate, favipiravir, or PBS and distilled water. PBS served as a control for intranasal administration; distilled water served as a control for oral administration. Body weights were monitored daily (a–c). Three mice per group were euthanized at 3 and 6 d.p.i., and the virus titres in lungs were determined by plaque assays in MDCK cells (d–f). Statistically significant

differences between virus titres of control mice and those of mice treated with antiviral drugs were determined by using Welch's *t*-test or Student's *t*-test on the result of the F-test. The resulting *P* values were corrected by using Holm's method (**P* < 0.05). **Virus was not recovered from all three animals infected with CA04 virus (for the 10^4 p.f.u. infection groups, the virus titres for the individual animals were $10^{2.0}$ p.f.u. ml^{-1} at 3 d.p.i. and $10^{2.2}$ and $10^{3.1}$ p.f.u. ml^{-1} at 6 d.p.i. For the 10^3 p.f.u. infection groups, the virus titres for the individual animals were $10^{2.2}$ and $10^{1.8}$ p.f.u. ml^{-1} at 3 d.p.i.). Error bars denote standard deviations.

bronchial epithelial cells²² (Fig. 3). This binding pattern may be influenced by the 'human-type' residues at position 226 (Anhui/1 possesses HA-226L and Hangzhou/1 possesses HA-226I, the residues also found in human H3N2 influenza viruses at this position). Shanghai/1 HA (encoding the 'avian-type' HA-226Q) was less selective, binding equally well to both α 2,6- and α 2,3-linked sialosides (Fig. 3). The specificity of the recombinant Anhui/1 and Hangzhou/1 HA viruses was comparable to the 'human-type' receptor specificity of H5N1 virus receptor mutants that exhibit respiratory droplet transmission in ferrets^{15,23}, and to that of the human H7N3 A/New York/107/2003 isolate that exhibits contact, but not respiratory droplet transmission, in ferrets²⁴. Notably, however, inhibition of the NA enzymatic function by inclusion of the neuraminidase inhibitor zanamivir in the glycan array analysis resulted in substantial increases in binding to α 2,3-linked sialosides (as shown for Hangzhou/1 in Supplementary Fig. 10). Moreover, preliminary analysis of the specificity of the recombinant Anhui/1 H7 HA revealed preferential binding to α 2,3 sialosides (R.P.d.V., R.M. & J.C.P., unpublished observations). Thus, the 'human-type' receptor specificity of A(H7N9) viruses assessed by glycan array seems to reflect the combined activities of HA and NA.

In pigs, we observed a mild infection with no clinical symptoms (Supplementary Tables 10 and 11, Supplementary Fig. 11 and Supplementary Information). Signs of disease were also absent in infected chickens and quails, which supported virus replication in a limited number of organs (a characteristic of low pathogenic avian influenza viruses) (Supplementary Tables 12–15 and Supplementary Information).

Antiviral compounds are currently the only therapeutic and prophylactic option for A(H7N9) infections. We therefore determined the *in vitro* 50% inhibitory concentration (IC_{50}) of several NA inhibitors (oseltamivir, zanamivir, laninamivir and peramivir), and of an experimental inhibitor of the viral RNA polymerase (favipiravir, also known as T-705) against egg-grown virus stocks of Anhui/1 and Shanghai/1. Sanger sequencing of the Shanghai/1 NA gene revealed the R294K mutation known to confer resistance to NA inhibitors for N2 and N9 NAs²⁵. Both Anhui/1 and Shanghai/1 were sensitive to all NA inhibitors tested (Supplementary Table 16), consistent with a recent report that Shanghai/1 is susceptible to oseltamivir and zanamivir²⁶, but inconsistent with the presence of the R294K mutation. A possible explanation for this discrepancy is that the Shanghai/1 isolate contained a mixture of drug-sensitive and -resistant NA genes in which the drug-resistant subpopulation may not have been detected, as described in ref. 27. In fact, plaque purification of the egg-grown Shanghai/1 virus stock revealed a mixed population of NA genes encoding NA-294R or -294K. Testing of these variants confirmed that Shanghai/1-NA-294R was susceptible to NA inhibitors, whereas the NA-294K variant was not (Supplementary Table 16). Further testing demonstrated that 30% of oseltamivir-sensitive virus in the mixture is sufficient for results to be consistent with a purely sensitive virus population (Supplementary Table 17). The IC_{50} values of favipiravir, determined by plaque-reduction assays in MDCK cells, were low for Anhui/1 and the control CA04 virus ($1.4 \mu\text{g ml}^{-1}$ and $1.2 \mu\text{g ml}^{-1}$,

respectively), suggesting that this compound could be a treatment option against A(H7N9) viruses resistant to NA inhibitors.

We also evaluated the therapeutic efficacy of the anti-influenza drugs in mice infected with Anhui/1, CA04 or a recombinant virus possessing the Shanghai/1 NA gene (encoding NA-294K) with the remaining genes from Anhui/1. Peramivir, which is structurally similar to oseltamivir but is administered intravenously, was omitted from these experiments. Mice infected with 10^3 and 10^4 p.f.u. of viruses were treated with the drugs beginning at 2 h post-infection. Some NA inhibitors had modest effects on the body-weight loss of Anhui/1-infected mice (Fig. 4a–c), consistent with limited, although statistically significant, effects on virus titre reduction (Fig. 4d–f). We currently do not know whether neuraminidase-resistant variants arose during Anhui/1 replication in mice, which may have limited virus susceptibility to NA inhibitors. In this context, it is interesting to note the poor efficacy of oseltamivir when used to treat a person infected with A(H7N9)²⁸, presumably owing to the emergence of drug-resistant variants. By contrast, favipiravir, which targets the viral polymerase complex, showed clear therapeutic effectiveness against both viruses at both doses tested.

In summary, on the basis of their sequences and phylogenetic relationships, Anhui/1 and Shanghai/1 originated from an avian host but possess several characteristic features of human influenza viruses, such as efficient binding to human-type receptors, efficient replication in mammalian cells (probably conferred by PB2-627K) and respiratory droplet transmission in ferrets (Anhui/1). These properties, together with the low efficacy of NA inhibitors and the lack of human immunity (we tested 500 human sera collected from various age groups in Japan and found no antibodies that recognized Anhui/1), make A(H7N9) viruses a formidable threat to public health.

METHODS SUMMARY

Viruses. A(H7N9) viruses were propagated in embryonated chicken eggs to produce viral stocks. Control viruses were propagated as described in Methods. All experiments with A(H7N9) viruses were carried out in approved biosafety level (BSL3) containment laboratories.

Animals. All animals were used according to approved protocols for their care and use. Detailed procedures are provided in Methods.

Antiviral sensitivity of viruses in mice. Six-week-old female BALB/c mice (Japan SLC; groups of six) were intranasally inoculated with 10^3 or 10^4 p.f.u. of Anhui/1, CA04 or a recombinant virus possessing the Shanghai/1 NA gene encoding NA-294K and the remaining genes from Anhui/1. At 2 h post-infection, mice were administered antiviral compounds as described in detail in Methods. Three mice per group were euthanized at 3 or 6 d.p.i. and the virus titres in lungs were determined by plaque assays in MDCK cells.

Full Methods and any associated references are available in the online version of the paper.

Received 23 May; accepted 20 June 2013.

Published online 10 July 2013.

- World Health Organization. China–WHO Joint Mission on Human Infection with Avian Influenza A(H7N9) Virus. 18–24 April 2013, Mission Report. http://www.who.int/influenza/human_animal_interface/influenza_h7n9/ChinaH7N9JointMissionReport2013.pdf (2013).
- Gao, R. *et al.* Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.* **368**, 1888–1897 (2013).
- World Health Organization. Number of confirmed human cases of avian influenza A(H7N9) reported to WHO. Report 8 – data in WHO/HQ as of 30 May 2013, 15:45 GMT +1. http://www.who.int/influenza/human_animal_interface/influenza_h7n9/08_ReportWebH7N9Number.pdf (2013).
- Kageyama, T. *et al.* Genetic analysis of novel avian A(H7N9) influenza viruses isolated from patients in China, February to April 2013. *Euro Surveill.* **18**, 20453 (2013).
- Liu, Q. *et al.* Genomic signature and protein sequence analysis of a novel influenza A (H7N9) virus that causes an outbreak in humans in China. *Microbes Infect.* (2013).
- Rogers, G. N. *et al.* Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature* **304**, 76–78 (1983).
- Hatta, M., Gao, P., Halfmann, P. & Kawaoka, Y. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science* **293**, 1840–1842 (2001).
- Subbarao, E. K., Kawaoka, Y. & Murphy, B. R. Rescue of an influenza A virus wild-type PB2 gene and a mutant derivative bearing a site-specific temperature-sensitive and attenuating mutation. *J. Virol.* **67**, 7223–7228 (1993).

- Li, Z. *et al.* Molecular basis of replication of duck H5N1 influenza viruses in a mammalian mouse model. *J. Virol.* **79**, 12058–12064 (2005).
- Itoh, Y. *et al.* In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses. *Nature* **460**, 1021–1025 (2009).
- van den Brand, J. M. *et al.* Severity of pneumonia due to new H1N1 influenza virus in ferrets is intermediate between that due to seasonal H1N1 virus and highly pathogenic avian influenza H5N1 virus. *J. Infect. Dis.* **201**, 993–999 (2010).
- Min, J. Y. *et al.* Classical swine H1N1 influenza viruses confer cross protection from swine-origin 2009 pandemic H1N1 influenza virus infection in mice and ferrets. *Virology* **408**, 128–133 (2010).
- Maines, T. R. *et al.* Transmission and pathogenesis of swine-origin 2009 A(H1N1) influenza viruses in ferrets and mice. *Science* **325**, 484–487 (2009).
- Munster, V. J. *et al.* Pathogenesis and transmission of swine-origin 2009 A(H1N1) influenza virus in ferrets. *Science* **325**, 481–483 (2009).
- Imai, M. *et al.* Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* **486**, 420–428 (2012).
- Jackson, S. *et al.* Reassortment between avian H5N1 and human H3N2 influenza viruses in ferrets: a public health risk assessment. *J. Virol.* **83**, 8131–8140 (2009).
- Maines, T. R. *et al.* Lack of transmission of H5N1 avian-human reassortant influenza viruses in a ferret model. *Proc. Natl Acad. Sci. USA* **103**, 12121–12126 (2006).
- Herfst, S. *et al.* Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* **336**, 1534–1541 (2012).
- Yamada, S. *et al.* Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature* **444**, 378–382 (2006).
- Chutinimitkul, S. *et al.* In vitro assessment of attachment pattern and replication efficiency of H5N1 influenza A viruses with altered receptor specificity. *J. Virol.* **84**, 6825–6833 (2010).
- Srinivasan, K., Raman, R., Jayaraman, A., Viswanathan, K. & Sasisekharan, R. Quantitative characterization of glycan-receptor binding of H9N2 influenza A virus hemagglutinin. *PLoS ONE* **8**, e59550 (2013).
- Chandrasekaran, A. *et al.* Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nature Biotechnol.* **26**, 107–113 (2008).
- Chen, L. M. *et al.* In vitro evolution of H5N1 avian influenza virus toward human-type receptor specificity. *Virology* **422**, 103–113 (2012).
- Belser, J. A. *et al.* Contemporary North American influenza H7 viruses possess human receptor specificity: implications for virus transmissibility. *Proc. Natl Acad. Sci. USA* **105**, 7558–7563 (2008).
- McKimm-Breschkin, J. L. *et al.* Mutations in a conserved residue in the influenza virus neuraminidase active site decreases sensitivity to Neu5Ac2en-derived inhibitors. *J. Virol.* **72**, 2456–2462 (1998).
- Centers for Disease Control & Prevention. Emergence of avian influenza A(H7N9) virus causing severe human illness - China, February–April 2013. *MMWR Morb. Mortal. Wkly Rep.* **62**, 366–371 (2013).
- Wetherall, N. T. *et al.* Evaluation of neuraminidase enzyme assays using different substrates to measure susceptibility of influenza virus clinical isolates to neuraminidase inhibitors: report of the neuraminidase inhibitor susceptibility network. *J. Clin. Microbiol.* **41**, 742–750 (2003).
- Liu, X. *et al.* Poor responses to oseltamivir treatment in a patient with influenza A (H7N9) virus infection. *Emerg. Microbes & Infections* **2**, e27 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank Y. Shu for A/Anhui/1/2013 (H7N9) and A/Shanghai/1/2013 (H7N9) viruses. We thank the IMSUT serum bank for providing human sera. We thank R. Webster for providing monoclonal antibody to A/seal/Massachusetts/1/80 (H7N7). Polyclonal anti-influenza virus H7 HA, A/Netherlands/219/2003 (H7N7) (anti-serum, goat) NR-9226, was obtained through the National Institutes of Health (NIH) Biodefense and Emerging Infections Research Resources Repository, National Institute of Allergy and Infectious Diseases (NIAID), NIH. We thank S. Watson for editing the manuscript. T. Suzuki, K. Takahashi, S. Fujisaki and H. Xu for discussions, and Y. Sato, H. Sugawara, A. Sato, M. Ejima and T. Miura for technical assistance. We thank Toyama Chemical Co. for providing favipiravir, Daiichi Sankyo Co. for providing laninamivir, F. Hoffmann–La Roche for providing oseltamivir carboxylate, GlaxoSmithKline for providing zanamivir and Shionogi & Co. for providing peramivir. This work was supported by the Japan Initiative for Global Research Network on Infectious Diseases from the Ministry of Education, Culture, Sports, Science and Technology, Japan; by grants-in-aid from the Ministry of Health, Labour and Welfare, Japan; by ERATO (Japan Science and Technology Agency); by NIAID Public Health Service research grants AI099274 and AI058113 to J.C.P., and by an NIAID-funded Center for Research on Influenza Pathogenesis (CRIP, HHSN266200700010C) to Y.K.

Author Contributions T.W., M.K., S. Fukuyama, M. Imai, S. Yamada, S.M., S. Yamayoshi, K.I.-H., Y. Sakoda, E.T., M.H., S.W., E.A.M., G.N., H.K., T.O., J.C.P., M.T. and Y.K. designed the study; T.W., M.K., S. Fukuyama, N.N., M. Imai, S. Yamada, S.M., S. Yamayoshi, K.I.-H., Y. Sakoda, E.T., R.M., T.N., M.H., H.J., D.Z., N.K., M.S., R.P.d.V., S.S., M. Okamatsu, T.T., Y.T., N.F., K.G., H.K., I.I., M. Ito, Y.S.-T., Y. Sugita, R.U., R.Y., A.J.E., G.Z., S. Fang, J.P., A.H., Y.U., T.S. and H.H. performed the experiments; T.W., M.K., S. Fang, N.N., M. Imai, S. Yamayoshi, S.M., S. Yamada, K.I.-H., Y. Sakoda, E.T., R.M., T.N., M.H., H.J., D.Z., R.P.V., S.S., T.T., Y.T., H.K., E.K. and H. H. analysed the data; T.W., S. Fukuyama, N.N., E.T., R.M., M.H., R.P.d.V., M. Ozawa, G.N., T.O., J.C.P., H.H., M.T. and Y.K. wrote the manuscript; Y.K. oversaw the project. T.W., M.K., S. Fukuyama, N.N. and M. Imai contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Y.K. (kawaokay@svm.vetmed.wisc.edu).

METHODS

Viruses. Anhui/1 and Shanghai/1, both kindly provided by Y. Shu, and Dk/GM466 were propagated in embryonated chicken eggs. For antigenic characterization by haemagglutination inhibition assays (see below for detailed procedures), Anhui/1 was also propagated in MDCK cells. CA04, Kawasaki/173 and Vietnam/1203 were propagated in MDCK cells. All experiments with H7N9 viruses were performed in enhanced biosafety level 3 (BSL3) containment laboratories at the University of Tokyo and the National Institute of Infectious Diseases, Japan, which are approved for such use by the Ministry of Agriculture, Forestry and Fisheries, Japan; or in enhanced BSL3 containment laboratories at the University of Wisconsin-Madison, which are approved for such use by the Centers for Disease Control and Prevention and by the US Department of Agriculture.

Cells. MDCK cells were maintained in Eagle's MEM containing 5% newborn calf serum. Human embryonic kidney 293T cells were maintained in DMEM containing 10% FCS. Normal human bronchial epithelial cells (NHBEs) were obtained from Lonza. The monolayers of NHBE cells were cultured and differentiated as previously described²⁹. All cells were incubated at 37 °C with 5% CO₂.

Antiviral compounds. Laninamivir and laninamivir octanoate were kindly provided by Daiichi Sankyo Co. Favipiravir was kindly provided by Toyama Chemical Co. and oseltamivir carboxylate was provided by F. Hoffmann-La Roche. Zanamivir was kindly provided by GlaxoSmithKline. Peramivir was kindly provided by Shionogi & Co.

Reverse genetics. Plasmid-based reverse genetics for influenza virus generation was performed as previously described³⁰. In brief, plasmids encoding the complementary DNAs for the eight viral RNA segments under the control of the human RNA polymerase I promoter and the mouse RNA polymerase I terminator (referred to as PolI plasmids), and plasmids for the expression of the viral PB2, PB1, PA and nucleoprotein proteins derived from a laboratory-adapted influenza A virus strain A/WSN/33 (H1N1), under the control of the chicken β -actin promoter³¹, were transfected into 293T cells with the help of a transfection reagent, Trans-IT 293 (Mirus). At 48 h post-transfection, culture supernatants were collected and inoculated to embryonated chicken eggs for virus propagation.

Growth kinetics of virus in cell culture. MDCK cells were infected in triplicate in 12-well plates with Anhui/1, Dk/GM466 or CA04 at a multiplicity of infection (m.o.i.) of 0.01. After incubation at 37 °C for 1 h, the viral inoculum was replaced with MEM containing 0.3% bovine serum albumin (with 0.75 μ g ml⁻¹ trypsin treated with L-1-tosylamide-2-phenylethyl chloromethyl ketone), followed by further incubation at 37 °C. Culture supernatants collected at the indicated time points were subjected to virus titration by using plaque assays in MDCK cells.

Cultures of differentiated NHBE cells grown on semipermeable membrane supports were washed extensively with DMEM to remove accumulated mucus and infected in triplicate with virus at a m.o.i. of 0.001 from the apical surface. The inoculum was removed after 30 min of incubation at 33 °C or 37 °C, and cells were further incubated at 33 °C or 37 °C. Samples were collected at 6, 12, 24, 48, 72 and 96 h post-infection from the apical surface. Apical collection was performed by adding 500 μ l of medium to the apical surface, followed by incubation for 30 min at 33 °C or 37 °C, and removal of the medium from the apical surface. The titres of viruses released into the cell culture supernatant were determined by plaque assay in MDCK cells.

Animal experiments. The sample sizes ($n = 3$) for the mouse, ferret, quail and chicken studies were chosen because they have previously been shown to be sufficient to evaluate a significant difference among groups^{10,14,32–34}. For the non-human primate and pig experiments, two or three animals per group were used and no statistical analysis was performed. No method of randomization was used to determine how animals were allocated to the experimental groups and processed in this study. The investigator was not blinded to the group allocation during the experiments or when assessing the outcome.

Experimental infection of mice. Six-week-old female BALB/c mice (Japan SLC) were used in this study. Baseline body weights were measured before infection. Under anaesthesia, four mice per group were intranasally inoculated with 10¹, 10², 10³, 10⁴, 10⁵ or 10⁶ p.f.u. (50 μ l) of Anhui/1, Shanghai/1, Dk/GM466 or CA04. Body weight and survival were monitored daily for 14 days. For virological and pathological examinations, six mice per group were intranasally infected with 10⁴ or 10⁶ p.f.u. (50 μ l) of the viruses and three mice per group were euthanized at 3 and 6 d.p.i. The virus titres in various organs were determined by plaque assays in MDCK cells. All experiments with mice were performed in accordance with the University of Tokyo's Regulations for Animal Care and Use and approved by the Animal Experiment Committee of the Institute of Medical Science, the University of Tokyo.

Experimental infection of ferrets. Five- to eight-month-old female ferrets (Triple F Farms), which were serologically negative by haemagglutination inhibition assay for currently circulating human influenza viruses, were used in this study. Under anaesthesia, six ferrets per group were intranasally inoculated with 10⁶ p.f.u.

(0.5 ml) of Anhui/1, Shanghai/1, Dk/GM466 or CA04. Three ferrets per group were euthanized at 3 and 6 d.p.i. for virological and pathological examinations. The virus titres in various organs were determined by plaque assays in MDCK cells. All experiments with ferrets were performed in accordance with the University of Tokyo's Regulations for Animal Care and Use and approved by the Animal Experiment Committee of the Institute of Medical Science, the University of Tokyo.

Ferret transmission study. Pairs of ferrets were individually housed in adjacent wireframe cages that prevented direct and indirect contact between animals but allowed spread of influenza virus by respiratory droplets. Under anaesthesia, three ferrets per group were intranasally inoculated with 0.5 ml of 10⁶ p.f.u. ml⁻¹ of Anhui/1, Dk/GM466 or CA04 (inoculated ferrets). One day after infection, three naive ferrets (contact ferrets) were each placed in a cage adjacent to an infected ferret (in these cages, infected and contact ferrets are separated by ~5 cm). Body weight and temperature were monitored every other day. Nasal washes were collected from infected ferrets on day 1 after inoculation and from contact ferrets on day 1 after co-housing, and then every other day (for up to 9 days) for virological examinations. The virus titres in nasal washes were determined by plaque assays in MDCK cells.

Experimental infection of cynomolgus macaques. Approximately 2-year-old male cynomolgus macaques (*Macaca fascicularis*) from Cambodia (obtained from Japan Laboratory Animals), weighing 2.2–2.8 kg and serologically negative by AniGen AIV antibody ELISA, which detects infection of all influenza A virus subtypes (Animal Genetics) and neutralization against A/Osaka/1365/2009 (H1N1pdm09), A/Kawasaki/UTK-4/2009 (seasonal H1N1), A/Kawasaki/UTK-20/2008 (H3N2), B/Tokyo/UT-E2/2008 (type B) and A/duck/Hong Kong/301/78 (H7N2) viruses, were used in this study. Under anaesthesia, six and four macaques were inoculated with Anhui/1 or Dk/GM466 (10⁷ p.f.u. ml⁻¹ each), respectively, through a combination of intratracheal (4.5 ml), intranasal (0.5 ml per nostril), ocular (0.1 ml per eye) and oral (1 ml) routes (resulting in a total infectious dose of 6.7×10^7 p.f.u.). Body temperature was monitored at 0, 1, 3, 5 and 6 d.p.i. by rectal thermometer. Nasal and tracheal swabs were collected at 1, 3, 5 and 6 d.p.i. for virological examinations. Three Anhui/1- and two Dk/GM466-infected macaques per group were euthanized at 3 and 6 d.p.i. for virological and pathological examinations. Virus titres were determined by plaque assays in MDCK cells. All experiments with macaques were performed in accordance with the Regulation on Animal Experimentation Guidelines at Kyoto University (5 February, 2007) and were approved by the Committee for Experimental Use of Nonhuman Primates in the Institute for Virus Research, Kyoto University.

Experimental infection of miniature pigs. Two- to three-month-old female specific-pathogen-free miniature pigs (NIBS line; Nippon Institute for Biological Science), which were serologically negative by neutralization assay for currently circulating human and swine influenza viruses, were used in this study. Baseline body temperatures were measured before infection. Four and two pigs were intranasally inoculated with 10⁷ p.f.u. (1 ml) of Anhui/1 or Dk/GM466, respectively. Body temperature was monitored daily. Nasal swabs were collected every day for virological examinations. Two pigs per group were euthanized at 3 d.p.i. for virological and pathological examinations; the remaining two Anhui/1-inoculated pigs were euthanized at 6 d.p.i. Virus titres were determined by plaque assays in MDCK cells. All experiments with miniature pigs were performed in accordance with guidelines established by the Animal Experiment Committee of the Graduate School of Veterinary Medicine, Hokkaido University, and were approved by the Institutional Animal Care and Use Committee of Hokkaido University.

Experimental infection of chickens. Four-week-old female specific-pathogen-free chickens (Nisseiken Co.) were used in this study. Six chickens per group were intranasally inoculated with 2×10^6 p.f.u. (0.2 ml) of Anhui/1 or Dk/GM466. Tracheal and cloacal swabs were collected every day for virological examinations. Three chickens per group were euthanized at 3 and 6 d.p.i. for virological examinations. The virus titres in various organs and swabs were determined by plaque assays in MDCK cells. All experiments with chickens were performed in accordance with the Animal Experimentation Guidelines of the National Institute of Infectious Disease (NIID) and were approved by the Animal Care and Use Committee of the NIID.

Experimental infection of quails. Four-month-old female quails (Yamanaka Koucho En) were used in this study. Six quails per group were intranasally inoculated with 2×10^6 p.f.u. (0.2 ml) of Anhui/1 or Dk/GM466. Tracheal and cloacal swabs were collected every day for virological examinations. Three quails per group were euthanized at 3 and 6 d.p.i. for virological examinations. The virus titres in various organs and swabs were determined by plaque assays in MDCK cells. All experiments with quails were performed in accordance with the University of Tokyo's Regulations for Animal Care and Use and were approved by the Animal Experiment Committee of the Institute of Medical Science, the University of Tokyo.

Pathological examination. Excised tissues of animal organs preserved in 10% phosphate-buffered formalin were processed for paraffin embedding and cut into 3- μ m-thick sections. One section from each tissue sample was stained using a standard haematoxylin and eosin procedure, whereas another one was processed for immunohistological staining with a rabbit polyclonal antibody for type A influenza nucleoprotein antigen (prepared in our laboratory) that reacts comparably with all of the viruses tested in this study. Specific antigen-antibody reactions were visualized with 3,3'-diaminobenzidine tetrahydrochloride staining by using the DAKO LSAB2 system (DAKO Cytomation).

Cytokine and chemokine measurement. Macaque lung homogenates and serum samples were processed with the MILLIPLEX MAP Non-human Primate Cytokine/Chemokine Panel-Premixed 23-Plex (Merck Millipore). Array analysis was performed by using the Bio-Plex Protein Array system (Bio-Rad Laboratories).

NA inhibition assay. *In vitro* NA activity of viruses was determined by using the commercially available NA-Fluor Influenza Neuraminidase Assay Kit (Applied Biosystems). In brief, diluted viruses were mixed with the indicated amounts of oseltamivir carboxylate, zanamivir, laninamivir or peramivir and incubated at 37 °C for 30 min. Methylumbelliferyl-*N*-acetylneuraminic acid (MUNANA) was then added as a fluorescent substrate, and the mixture was incubated at 37 °C for 1 h. The reaction was stopped by adding 0.12 M Na₂CO₃ in 40% ethanol. The fluorescence of the solution was measured at an excitation wavelength of 355 nm and an emission wavelength of 460 nm, and the IC₅₀ values were calculated.

Polymerase inhibitor sensitivity assay. 8 × 10⁵ MDCK cells were infected with approximately 50 p.f.u. of viruses. After incubation at 37 °C for 1 h, the viral inoculum was replaced with agarose medium containing various concentrations of favipiravir. After the cells were incubated at 37 °C for 2 days, plaques were visualized by crystal violet staining and counted.

Antiviral sensitivity of viruses in mice. Under anaesthesia, six mice per group were intranasally inoculated with 10³ or 10⁴ p.f.u. (50 μ l) of Anhui/1, CA04 or a recombinant virus possessing the Shanghai/1 NA gene encoding NA-294K and the remaining genes from Anhui/1. At 2 h after inoculation, mice were treated with the following antiviral compounds: (1) oseltamivir phosphate: 4 or 40 mg per kg per 200 μ l, administered orally twice a day for 5 days; (2) zanamivir: 0.8 or 8 mg per kg per 50 μ l, administered intranasally once daily for 5 days; (3) laninamivir: 0.75 mg per kg per 50 μ l, administered intranasally once during the entire experimental course; (4) favipiravir: 30 or 150 mg per kg per 200 μ l, administered orally twice a day for 5 days; (5) or PBS intranasally (50 μ l) and distilled water orally administered. For virological examinations, three mice per group were euthanized at 3 and 6 d.p.i. The virus titres in lungs were determined by plaque assays in MDCK cells.

Antigenicity characterization by HI assays. Anti-H7 HA monoclonal antibodies 46/6, 46/2 and 55/3 against A/seal/Massachusetts/1/80 (H7N7) virus were kindly provided by R. G. Webster. The goat polyclonal antibody NR-9226 (raised against A/Netherlands/219/2003 (H7N7)) was obtained from BEI Resources. The remaining antibodies, that is, mouse monoclonal antibodies B1275m and B1275m (raised against A/Netherlands/219/2003 (H7N7), MyBioSource), 127-10023 (raised against A/FPV/Rostock/34 (H7N1), RayBiotech), 10H9, 9A9 and 1H11 (raised against A/FPV/Rostock/34 (H7N1), HyTest), and rabbit polyclonal antibody MBS432028 (raised against A/chicken/MD/MINHMA/2004 (H7N2), MyBioSource) were commercially available. Anhui/1 propagated in embryonated chicken eggs or in MDCK cells was used in this study. Antibodies were serially diluted twofold with PBS in 96-well U-bottom microtitre plates and mixed with the amount of virus equivalent to eight haemagglutination units, followed by incubation at room temperature (25 °C) for 30 min. After adding 50 μ l of 0.5% turkey red blood cells, the mixtures were gently mixed and incubated at room temperature for a further 45 min. haemagglutination inhibition titres are expressed as the inverse of the highest antibody dilution that inhibited haemagglutination (Supplementary Table 18). These data were used to select antibodies for glycan arrays.

Serology with human sera. Human sera, collected in Japan in November 2012 from 200 donors ranging in age from 20 to 63 years, were treated with receptor-destroying enzyme (Denka Seiken Co). Twofold serial dilutions of the treated sera were mixed with 100 p.f.u. of Anhui/1 and incubated at 37 °C for 1 h. MDCK cells were inoculated with the virus-serum mixtures and cultured for 3 days. The neutralizing activity of the sera was determined based on the cytopathic effects in inoculated cells. All experiments with human sera were approved by the Research Ethics Review Committee of the Institute of Medical Science, the University of Tokyo (approval number: 21-38-1117).

A total of 300 serum samples were also obtained from the serum bank of the National Institute of Infectious Diseases in Japan. Samples were collected from

different regions of Japan during 2010–2011. Subjects were divided into 10 age groups, 30 samples per group, and analysed for antibodies against Anhui/1 by use of the haemagglutination inhibition assay with 0.5% turkey red blood cells.

Glycan arrays. Glycan array analysis was performed on a glass slide microarray containing 6 replicates of 57 diverse sialic acid-containing glycans including terminal sequences and intact *N*-linked and *O*-linked glycans found on mammalian and avian glycoproteins and glycolipids³⁵. β -propiolactone-inactivated viruses containing A(H7N9) virus HA and NA genes in the background of A/Puerto Rico/8/34 (H1N1) virus were applied to the array at dilutions of 128–512 haemagglutination units ml⁻¹. After incubation at room temperature for 1 h, slides were washed and overlaid with a 1:200 dilution of rabbit or goat anti-H7 antibody for 1 h (selected on the basis of antibody characterization with several H7 viruses; see Supplementary Table 18), and finally with anti-rabbit IgG Alexa Fluor 488 or anti-goat IgG Alexa Fluor 647 (Invitrogen) at 10 μ g ml⁻¹. Slides were then washed and scanned on a ProScanArray Express HT (PerkinElmer) confocal slide scanner to detect bound virus. A complete list of glycans present on the array is provided in Supplementary Materials (Supplementary Table 9).

Electron microscopy. Anhui/1 was inoculated into 10-day-old embryonated chicken eggs and the allantoic membranes were collected 24 h after inoculation. They were then processed for ultrathin section electron microscopy and scanning electron microscopy as described previously^{10,36}.

Statistical analysis. All statistical analyses were performed using JMP Pro 9.0.2 (SAS Institute). Statistically significant differences between the virus titres of Dk/GM466-infected mice and those of other mice were determined by using Welch's *t*-test with Bonferroni's correction. Comparisons of virus titres in antiviral sensitivity assays in mice were also done using Welch's *t*-test or Student's *t*-test on the result of the F-test. The resulting *P* values were corrected by using the Holm's method.

Biosafety and biosecurity. All recombinant DNA protocols were approved by the University of Wisconsin-Madison's Institutional Biosafety Committee after risk assessments were conducted by the Office of Biological Safety, and by the University of Tokyo's Subcommittee on Living Modified Organisms, and, when required, by the competent minister of Japan. In addition, the University of Wisconsin-Madison Biosecurity Task Force regularly reviews the research program and ongoing activities of the laboratory. The task force has a diverse skill set and provides support in the areas of biosafety, facilities, compliance, security and health. Members of the Biosecurity Task Force are in frequent contact with the principal investigator and laboratory personnel to provide oversight and assure biosecurity. All experiments with H7N9 viruses were performed in enhanced BSL3 containment laboratories. Ferret transmission studies were conducted by two scientists with DVM and/or PhD degrees that each had more than 5 years of experience working with highly pathogenic influenza viruses and performing animal studies with such viruses. Our staff who work with ferrets, nonhuman primates, pigs, chickens and quails wear disposable overalls and powered air-purifying respirators that filter the air. Biosecurity monitoring of the facilities is ongoing. All personnel complete rigorous biosafety, BSL3 and Select Agent (for the US laboratory) training before participating in BSL3-level experiments. The principal investigator participates in training sessions and emphasizes compliance to maintain safe operations and a responsible research environment. The laboratory occupational health plans are in compliance with policies at the respective institutions (University of Wisconsin-Madison and the University of Tokyo Occupational Health Programs).

29. Jakiela, B., Brockman-Schneider, R., Amineva, S., Lee, W. M. & Gern, J. E. Basal cells of differentiated bronchial epithelium are more susceptible to rhinovirus infection. *Am. J. Respir. Cell Mol. Biol.* **38**, 517–523 (2008).
30. Neumann, G. *et al.* Generation of influenza A viruses entirely from cloned cDNAs. *Proc. Natl Acad. Sci. USA* **96**, 9345–9350 (1999).
31. Niwa, H., Yamamura, K. & Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193–199 (1991).
32. Makarova, N. V., Ozaki, H., Kida, H., Webster, R. G. & Perez, D. R. Replication and transmission of influenza viruses in Japanese quail. *Virology* **310**, 8–15 (2003).
33. Liu, J. *et al.* Highly pathogenic H5N1 influenza virus infection in migratory birds. *Science* **309**, 1206 (2005).
34. Zhu, H. *et al.* Infectivity, transmission, and pathology of human H7N9 influenza in ferrets and pigs. *Science* <http://dx.doi.org/10.1126/science.1239844> (2013).
35. Xu, R. *et al.* A recurring motif for antibody recognition of the receptor-binding site of influenza hemagglutinin. *Nature Struct. Mol. Biol.* **20**, 363–370 (2013).
36. Noda, T. *et al.* Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature* **439**, 490–492 (2006).

Pathogenesis and transmission of avian influenza A (H7N9) virus in ferrets and mice

Jessica A. Belser¹, Kortney M. Gustin¹, Melissa B. Pearce¹, Taronna R. Maines¹, Hui Zeng¹, Claudia Pappas¹, Xiangjie Sun¹, Paul J. Carney¹, Julie M. Villanueva¹, James Stevens¹, Jacqueline M. Katz¹ & Terrence M. Tumpey¹

On 29 March 2013, the Chinese Center for Disease Control and Prevention confirmed the first reported case of human infection with an avian influenza A(H7N9) virus¹. The recent human infections with H7N9 virus, totalling over 130 cases with 39 fatalities to date, have been characterized by severe pulmonary disease and acute respiratory distress syndrome (ARDS)². This is concerning because H7 viruses have typically been associated with ocular disease in humans, rather than severe respiratory disease³. This recent outbreak underscores the need to better understand the pathogenesis and transmission of these viruses in mammals. Here we assess the ability of A/Anhui/1/2013 and A/Shanghai/1/2013 (H7N9) viruses, isolated from fatal human cases, to cause disease in mice and ferrets and to transmit to naive animals. Both H7N9 viruses replicated to higher titre in human airway epithelial cells and in the respiratory tract of ferrets compared to a seasonal H3N2 virus. Moreover, the H7N9 viruses showed greater infectivity and lethality in mice compared to genetically related H7N9 and H9N2 viruses. The H7N9 viruses were readily transmitted to naive ferrets through direct contact but, unlike the seasonal H3N2 virus, did not transmit readily by respiratory droplets. The lack of efficient respiratory droplet transmission was corroborated by low receptor-binding specificity for human-like α 2,6-linked sialosides. Our results indicate that H7N9 viruses have the capacity for efficient replication in mammals and human airway cells and highlight the need for continued public health surveillance of this emerging virus.

Laboratory-confirmed H7N9 infections in humans have been reported in eight contiguous provinces in eastern China (Anhui, Fujian, Henan, Hunan, Jiangsu, Jiangxi, Shandong and Zhejiang), two municipalities (Beijing and Shanghai) and Taiwan to date⁴. Although historically this subtype resulted in mild illness among humans³, the majority of novel A(H7N9) cases have resulted in severe respiratory illness in adults, representing the largest occurrence of H7 virus infections in humans². Unlike highly pathogenic avian influenza (HPAI) H5N1 viruses, which exhibit lethality in poultry and cause severe human disease, low pathogenic (LPAI) H7N9 viruses seem to be asymptomatic or cause mild disease in poultry and wild bird populations, possibly due to the presence of a single amino acid in the haemagglutinin (HA) cleavage site of this virus^{2,5}. Similar to H5N1 viruses, human cases with H7N9 virus seem to result from direct contact with infected poultry, yet >20% of confirmed H7N9 cases have reported no exposure to live animals⁶. Although sustained human-to-human transmission has not been observed with this virus to date, reports of family clusters of A(H7N9) infection suggest the potential for virus spread between close contacts⁶. Understanding the properties of A(H7N9) viruses which contribute to human disease and their capacity for human-to-human transmission is a critical requirement for guidance of public health responses. The ferret (*Mustela putorius furo*) is recognized as the best small mammalian model for the concurrent study of influenza virus pathogenicity and transmissibility, and was used here to assess the virulence of A(H7N9) viruses.

A/Anhui/1/13 (Anhui/1) virus was isolated from a throat-swab specimen collected from an adult female patient with known poultry exposure and who died from ARDS 6 days after illness onset. A/Shanghai/1/13 (Shanghai/1) virus was isolated from an adult male patient with no known poultry exposure who died from refractory hypoxemia 13 days after illness onset². These A(H7N9) viruses from China possess several molecular markers previously associated with human adaptation, including the presence of E627K in the PB2 protein and Q226L in the 210-loop in the HA gene². Glycan microarray analysis of these phylogenetically distinct viruses revealed that Shanghai/1 (bearing Q226 in the HA (H3 numbering)) largely binds to avian α 2,3 receptor analogues (glycans 4 to 40) (Fig. 1a and Supplementary Table 1), whereas Anhui/1 virus (bearing L226), revealed a mixed α 2,3/ α 2,6 receptor preference (Fig. 1b and Supplementary Table 1); the single Q226L amino substitution has long been recognized to modulate receptor-binding specificity⁷. In particular, Anhui/1 virus bound to human biantennary structures (glycans 43 to 47), as well as the longer, linear α 2,6 sialylated tri-*N*-acetyl lactosamine (glycans 58–60), some of which have been detected in *N* glycans of cultured human bronchial epithelial cells⁸. These findings are in accord with recent studies showing that Anhui/1 virus binds to human tracheal tissue sections, albeit at a lower intensity than other human-adapted viruses^{9,10}.

Select contemporary North American and Eurasian H7 viruses associated with human infection have previously demonstrated the capacity to transmit between ferrets in a direct-contact model, but not by respiratory droplets in the absence of direct contact¹¹. To characterize the transmissibility of A(H7N9) isolates, we intranasally inoculated ferrets with 10⁶ plaque-forming units (p.f.u.) of virus. Approximately 24 h post-inoculation (p.i.), inoculated–contact pairs were established by placing a naive ferret in the same cage as an inoculated ferret (direct contact transmission), or in an adjacent cage with perforated side-walls (respiratory droplet transmission)¹². Nasal washes were collected on alternate days from all animals and titred for the presence of infectious virus¹². The A(H7N9) viruses were compared to a seasonal H3N2 virus, A/Texas/50/2012 (Texas/50), representative of H3N2 viruses circulating in the 2012–2013 northern hemisphere season¹³.

Ferrets inoculated with Anhui/1 or Shanghai/1 virus showed modest weight loss and transient inactivity, but generally did not exhibit signs of severe disease (Table 1 and Supplementary Fig. 1). However, two of eight Shanghai/1 virus-inoculated ferrets were euthanized due to severe lethargy (laboured breathing, tremors, unresponsiveness day 4 p.i.) or excessive weight loss (26.6% drop from initial body weight day 12 p.i.) (Table 1 and Supplementary Fig. 1). A(H7N9) virus was detected at elevated titres in the nasal turbinates and trachea (Supplementary Fig. 2). However, unlike the seasonal H3N2 virus, A(H7N9) virus was detected in the lower respiratory tract (peak mean titre >10⁵ p.f.u. per g lung tissue) (Supplementary Fig. 1)^{10,14}. Infectious virus was detected in the brain and olfactory bulb of 1/3 and 2/3 ferrets inoculated with Anhui/1 and Shanghai/1 viruses, respectively (Supplementary Fig. 2); viraemia or

¹Influenza Division, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333, USA.

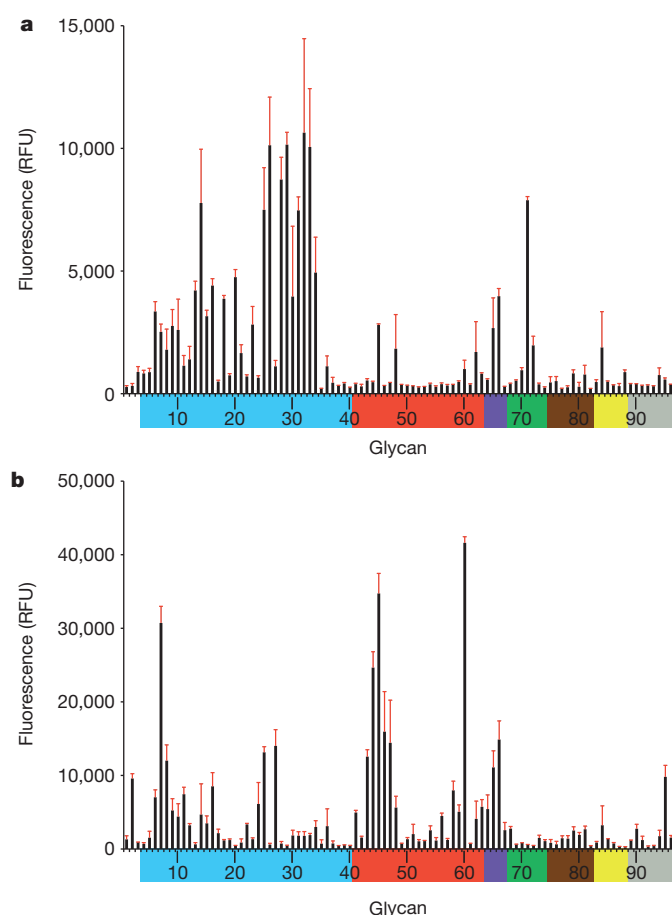


Figure 1 | Receptor specificity of A(H7N9) viruses. **a, b,** Glycan microarray analysis of Shanghai/1 (**a**) and Anhui/1 (**b**) viruses. Coloured bars highlight glycans that contain α 2,3 sialosides (blue), α 2,6 sialosides (red), α 2,6/ α 2,3 mixed sialosides (purple), N -glycolyl sialosides (green), α 2,8 sialosides (brown), β 2,6 and 9- O -acetyl sialosides (yellow), and non-sialoside glycans (grey) and are expressed as relative fluorescence units (RFU). Error bars reflect the s.e.m. in the signal for six independent replicates on the array. Structures of each of the numbered glycans are found in Supplementary Table 1.

systemic spread of virus to the spleen, kidney or liver was not detected. Similar to the 2009 H1N1 pandemic virus, gastrointestinal symptoms have been noted following A(H7N9) virus infection in humans²; rectal swabs were positive for virus detection in 7/11 and 5/8 Anhui/1 and Shanghai/1 virus-inoculated ferrets, respectively, whereas H3N2 virus was not detected in this sample (Supplementary Fig. 3). Furthermore, intestinal tract samples were positive in 3/3 Shanghai/1 inoculated ferrets day 3 p.i. (Supplementary Fig. 2). In agreement with reports of lymphopenia among A(H7N9) virus-infected patients⁵, ferrets showed transient lymphopenia that was sustained through day 7 p.i., with greater reductions in circulating lymphocytes in A(H7N9) virus-infected ferrets compared to ferrets infected with H3N2 virus (Supplementary Table 2).

Ferrets inoculated with A(H7N9) virus shed high titres of virus in nasal washes as early as day 1 p.i., significantly higher than Texas/50 virus-infected ferrets at this time point ($P < 0.05$) (Table 1 and Supplementary Fig. 4). Mean titres in nasal washes were sustained at $>10^5$ p.f.u. in Anhui/1 and Shanghai/1 virus-infected ferrets through day 3 p.i. (Supplementary Fig. 4). Efficient direct contact transmission occurred by day 1 post-contact, with viral titres in nasal wash samples reaching comparable titres to inoculated ferrets by day 3 post-contact. Respiratory droplet transmission of Anhui/1 virus was observed in 2/6 ferrets, with detection of virus in contact ferret nasal washes delayed 2–4 days compared to seasonal H3N2 virus; Shanghai/1 virus was detected in 1/3 ferrets in this model (Supplementary Fig. 4). These results are in accord with a recently published study demonstrating only limited transmissibility of A/Shanghai/2/13 virus by the airborne route¹⁵. Sequence analysis of virus gene segments obtained from nasal wash samples of A(H7N9) respiratory droplet contact animals (total $n = 3$) revealed no substantial amino acid changes. No sequence changes were observed from nasal washes obtained from the Shanghai/1 respiratory droplet contact ferret compared to Shanghai/1 sequences submitted to GISAID. Sequencing of nasal wash samples obtained on days 5–7 p.i. from Anhui/1 respiratory droplet contact ferrets revealed only one amino acid substitution in the HA protein at R140M (H3 numbering), not found among Anhui/1 inoculated ferrets. This position is located in the HA 140 loop region of HA1 and is not known to be associated with changes in H7 HA receptor binding or influenza virus transmission. Although A(H7N9) viruses are not readily transmissible by respiratory droplets, the efficient virus transmission observed in the direct-contact model has not been observed with H5N1 viruses¹² and may indicate the capacity of an A(H7N9)-like virus to acquire properties that would confer efficient airborne transmission.

Phylogenetic analysis showed that the H7N9 virus is a triple reassortant comprising an HA from an H7N3 virus, the neuraminidase (NA) of an H7N9 virus, and all six internal genes from an H9N2 virus². To assess the relative contribution of these avian precursors to the pathogenicity of the 2013 H7N9 viruses, we inoculated BALB/c mice intranasally with A/shoveler/Egypt/00215-NAMRU3/07 (shv/Egypt/07)¹⁶, an H7N9 virus with high surface glycoprotein amino acid sequence identity to Anhui/1, A/chicken/Vietnam/NCVD-1156/2011 (ck/VN/11), an H9N2 virus possessing internal genes with high genetic similarity to Anhui/1 virus, and the 2013 A(H7N9) viruses to determine virus replication, morbidity (as measured by weight loss), the 50% mouse infectious dose (MID_{50}), and the 50% mouse lethal dose (LD_{50}). MID_{50} titres were markedly low for all viruses, demonstrating high infectivity of the avian viruses in this model; typically human viruses do not replicate efficiently in mice without prior host adaptation¹⁷ (Table 2). Mice inoculated with Anhui/1 or Shanghai/1 virus showed severe morbidity ($>20\%$ weight loss) and a LD_{50} of $10^{3.3}$ – $10^{3.4}$, a value comparable to the select group of HPAI H5N1 and H7N7 viruses which have a high-pathogenicity phenotype in this model^{18,19}. In contrast, the avian precursor viruses shv/Egypt/07 and ck/VN/11 caused moderate morbidity in mice, but did not display a lethal phenotype (Table 2). Collectively, lung titres in Anhui/1 and Shanghai/1 virus-inoculated mice were significantly higher than LPAI precursor viruses examined day 3 p.i. ($P < 0.05$); by day 6 p.i., all H7N9 viruses were detected at comparable

Table 1 | Clinical signs, virus replication, and transmission of H7N9 and H3N2 viruses in ferrets

Virus	Inoculated animals					Direct contact animals		Respiratory droplet contact animals	
	Weight loss† (%)	Respiratory symptoms‡	Lethargy§	Lethality	Nasal wash titre¶	Virus detected	Seroconverted	Virus detected	Seroconverted
Anhui/1	8.8	3/11	1.1 (7/11)	0/11	7.1 (1)*	4/4	4/4	2/6	2/6
Shanghai/1	9.2	5/8	1.3 (7/8)	2/8	6.7 (1)*	4/4	4/4	1/3	1/3
Texas/50	10.1	4/6	1.0 (2/6)	0/6	5.6 (1)	3/3	3/3	3/3	3/3

† The percentage mean maximum weight loss observed during the first 10 days p.i.

‡ Number of ferrets that exhibited sneezing or nasal discharge during the first 10 days p.i.

§ Relative inactivity index of ferrets during the first 10 days p.i. (number of ferrets with observed lethargy in parentheses).

|| Number of animals euthanized before the end of the 14 day experimental period because of reaching a clinical end point.

¶ Peak mean \log_{10} (p.f.u. ml⁻¹) nasal wash titre (day p.i. in parentheses). * $P < 0.05$ compared with Texas/50 by one-way ANOVA.

Table 2 | Pathogenicity of H7N9 and related viruses in BALB/c mice

Virus	Subtype	Weight loss†	Lung titre (d 3)‡	Lung titre (d 6)‡	MID ₅₀ §	LD ₅₀ §
Anhui/1	H7N9	22.2 (7)	7.2 ± 0.6	5.5 ± 0.5	0.25	3.4
Shanghai/1	H7N9	25.6 (4)	6.6 ± 0.2	5.6 ± 0.2	0.25	3.3
shv/Egypt/07	H7N9	16.5 (3)	4.4 ± 0.2	5.6 ± 0.3	2.25	>6
ck/VN/11	H9N2	18.1 (6)	4.8 ± 0.2	3.1 ± 0.1	1.5	>6

† Maximum per cent weight loss (5 mice per group) after infection with 10⁶ p.f.u. of virus (day p.i. in parentheses).

‡ Average lung titres of three mice on indicated days p.i., expressed as log₁₀ (p.f.u. ml⁻¹) ± SD.

§ MID₅₀ and LD₅₀ are expressed as the log₁₀ p.f.u. ml⁻¹ required to give one MID₅₀ or one LD₅₀.

titre, >100-fold higher than the H9N2 virus at this time point ($P < 0.05$). All viruses examined replicated to low titres in the nose day 3 and 6 p.i. (Supplementary Fig. 5), and systemic spread of virus to the brain was not detected.

Unlike past human infections with H7 viruses, conjunctivitis has not been reported among A(H7N9)-infected individuals⁵. Viral titres in ferret eye or conjunctiva samples were not frequently detected in H7N9 or H3N2 virus-infected animals, although conjunctival washes from 50% of Anhui/1 virus-infected ferrets were positive for virus (Supplementary Figs 2 and 3). Ocular inoculation of mice with shv/Egypt, Anhui/1 or Shanghai/1 H7N9 viruses did not result in consistent virus replication in eye tissue (Supplementary Table 3). However, viral titres in the nose were detected days 3 and 6 p.i. following ocular inoculation with Anhui/1 and Shanghai/1 virus in mice with greater frequency compared to shv/Egypt/07 virus (Supplementary Table 3). These studies provide evidence that the 2013 A(H7N9) virus does not maintain the ocular tropism typical of H7 viruses, but is capable of using the eye as an entry portal to cause a productive respiratory infection.

The ability of Anhui/1 virus to form plaques in the presence or absence of the trypsin protease was assayed in MDCK cells by standard

plaque methods²⁰. Proteolytic cleavage of the HA molecule is a prerequisite for multi-cycle replication, and the ability of virus to replicate in the absence of trypsin is thought to be an important determinant of influenza virus pathogenicity in mammals²¹. Similar to the seasonal Texas/50 virus which requires an exogenous protease source for multi-cycle replication and plaque formation, the A(H7N9) virus failed to form visible plaques without the addition of trypsin (data not shown).

Because the airway epithelium is the primary site of influenza virus replication in humans, we investigated the A(H7N9) virus replication efficiency in Calu-3 cells, derived from human bronchial epithelium and grown on permeable cell culture membranes to resemble the *in vivo* airway epithelium²⁰ (Fig. 2a). For all viruses tested, titres of progeny virus progressively increased during the first 48 h p.i. Anhui/1 virus showed a 20- to 400-fold increase in replication at 48 h p.i. when compared to the other virus subtypes tested in Calu-3 cells ($P < 0.04$). Notably, compared to the seasonal H3N2 virus, Anhui/1 virus showed an 80,000-fold increase in replication at 24 h p.i.

As previous studies have suggested an association between the transmissibility of influenza viruses in ferrets and their ability to replicate efficiently at the lower temperature (33 °C), found in the environment of the mammalian upper airway²², we evaluated the replication kinetics of A(H7N9) virus in Calu-3 cells cultured at either 33 °C or 37 °C. Whereas the seasonal Texas/50 virus replicated equally well at both temperatures, A(H7N9) viruses showed significantly reduced replication at 33 °C compared with 37 °C at 24 h, characteristic of avian subtype viruses (Fig. 2b)²¹; comparable titres at either temperature were detected by 48 h for all viruses (data not shown). These data demonstrate that the A(H7N9) viruses replicate productively in polarized bronchial epithelial cells at 37 °C, but less efficiently at the lower temperature of the human proximal airways early after infection.

The recent cases of human infection with H7N9 viruses highlight the need to better understand the potential of these viruses to spread and cause disease in humans. Although the novel H7N9 influenza virus has caused severe illness and death among individuals in China, the initial epidemiologic findings suggest that there is no evidence of sustained human-to-human transmission of this virus. Our ferret transmission results are consistent with this observation and suggest that additional virus adaptation in mammals would be required to reach the high-transmissible phenotypes observed by the respiratory droplet route with pandemic and seasonal influenza A viruses¹⁰.

METHODS SUMMARY

Viruses and cells. H7N9 and H9N2 viruses were passaged in eggs; H1N1 virus was propagated in MDCK cells. Detailed culture conditions for Calu-3 cells are included in Methods. Virus titrations were performed in MDCK cells by standard plaque assay²⁰.

Biosafety and biosecurity. Experiments were conducted in biosafety level 3 facilities with enhancements (BSL3+) in accordance with U.S. Federal and WHO guidelines. Features of the BSL3+ containment laboratories include a double door entrance, double door autoclave, HEPA filtration of intake and exhaust air, shower out room, and a liquid effluent sterilizer. Security is monitored through an automated access control system and entrance to the laboratory is restricted by guard stations, cameras, card-controlled doors and biometric readers. Use of powered air-purifying respirators in the laboratory and yearly training requirements are mandatory for all staff with access.

Glycan arrays. Microarray printing and influenza virus analyses were performed as previously described¹¹; specific details on the glycans analysed are provided in Methods.

Animals. Six-to-eight week old female BALB/c mice (Jackson Laboratory) and six-to-seven month old influenza virus-seronegative male ferrets (Triple F Farms) were used in all experiments. All intranasally inoculated animals were administered 10⁶ p.f.u. of virus except for determination of mouse infectivity and lethality, where mice were administered serial 10-fold dilutions of virus. Virus transmission in ferrets was measured by co-housing an inoculated and naive ferret (direct contact transmission) or by placing an inoculated and naive ferret in adjacent cages with a perforated side wall (respiratory droplet transmission)¹². Transmission was determined by presence of virus in nasal washes and seroconversion

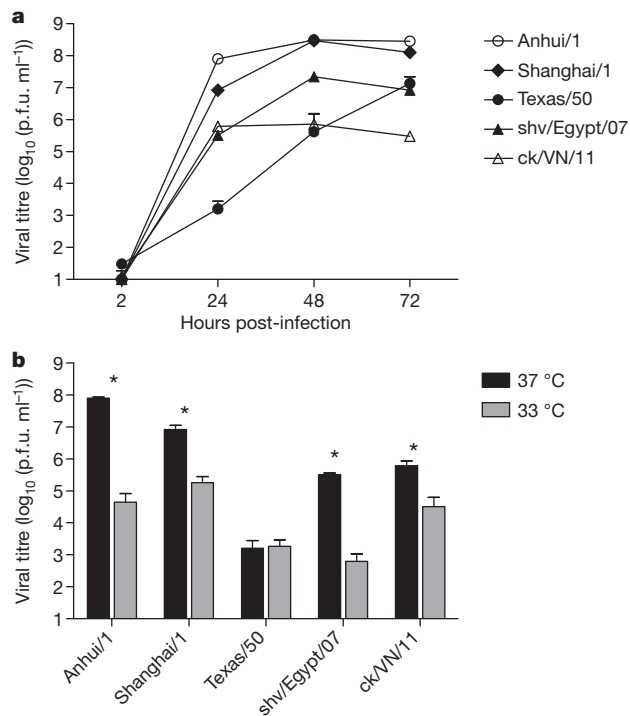


Figure 2 | Replication kinetics of A(H7N9) influenza viruses in polarized human airway epithelial cells. a, Replication at 37 °C; b, viral titre at 24 h post-infection. Calu-3 cells cultured for 1 week were inoculated with virus at an MOI of 0.01. Culture supernatants were collected at the times indicated and virus titres were determined by plaque assay. Titre values represent the average for three independent wells plus s.d. *indicates statistically significant difference between 37 °C and 33 °C for each virus, based on the Mann–Whitney test ($P < 0.004$).

of contact ferrets. Detailed procedures are included in Methods for pathotyping, transmission and tropism studies.

Full Methods and any associated references are available in the online version of the paper.

Received 18 May; accepted 20 June 2013.

Published online 10 July 2013.

- WHO. Avian influenza A(H7N9) virus http://www.who.int/influenza/human_animal_interface/influenza_h7n9/en/index.html (2013).
- Gao, R. *et al.* Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.* **368**, 1888–1897 (2013).
- Belser, J. A., Bridges, C. B., Katz, J. M. & Tumpey, T. M. Past, present, and possible future human infection with influenza virus A subtype H7. *Emerg. Infect. Dis.* **15**, 859–865 (2009).
- CDC. Emergence of avian influenza A(H7N9) virus causing severe human illness - China, February–April 2013. *MMWR Morb. Mortal. Wkly. Rep.* **62**, 1–4 (2013).
- Chen, Y. *et al.* Human infections with the emerging avian influenza A H7N9 virus from wet market poultry: clinical analysis and characterisation of viral genome. *Lancet* **381**, 1916–1925 (2013).
- Li, Q. *et al.* Preliminary report: epidemiology of the avian influenza A (H7N9) outbreak in China. *N. Engl. J. Med.* <http://dx.doi.org/10.1056/NEJMoa1304617> (2013).
- Rogers, G. N. *et al.* Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature* **304**, 76–78 (1983).
- Chandrasekaran, A. *et al.* Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nature Biotechnol.* **26**, 107–113 (2008).
- Tharakaraman, K. *et al.* Glycan receptor binding of the influenza A virus H7N9 hemagglutinin. *Cell* **153**, 1486–1493 (2013).
- Maines, T. R. *et al.* Transmission and pathogenesis of swine-origin 2009 A(H1N1) influenza viruses in ferrets and mice. *Science* **325**, 484–487 (2009).
- Belser, J. A. *et al.* Contemporary North American influenza H7 viruses possess human receptor specificity: Implications for virus transmissibility. *Proc. Natl Acad. Sci. USA* **105**, 7558–7563 (2008).
- Maines, T. R. *et al.* Lack of transmission of H5N1 avian-human reassortant influenza viruses in a ferret model. *Proc. Natl Acad. Sci. USA* **103**, 12121–12126 (2006).
- WHO. Recommended composition of influenza virus vaccines for use in the 2013–2014 northern hemisphere influenza season http://www.who.int/influenza/vaccines/virus/recommendations/201302_recommendation.pdf (2013).
- Zitzow, L. A. *et al.* Pathogenesis of avian influenza A (H5N1) viruses in ferrets. *J. Virol.* **76**, 4420–4429 (2002).
- Zhu, H. *et al.* Infectivity, transmission, and pathology of human H7N9 influenza in ferrets and pigs. *Science* <http://dx.doi.org/10.1126/science.1239844> (2013).
- Gerloff, N. A. *et al.* A high diversity of Eurasian lineage low pathogenicity avian influenza A viruses circulate among wild birds sampled in Egypt. *PLoS ONE* (in the press).
- Hartley, C. A., Reading, P. C., Ward, A. C. & Anders, E. M. Changes in the hemagglutinin molecule of influenza type A (H3N2) virus associated with increased virulence for mice. *Arch. Virol.* **142**, 75–88 (1997).
- Maines, T. R. *et al.* Avian influenza (H5N1) viruses isolated from humans in Asia in 2004 exhibit increased virulence in mammals. *J. Virol.* **79**, 11788–11800 (2005).
- de Wit, E. *et al.* Protection of mice against lethal infection with highly pathogenic H7N7 influenza A virus by using a recombinant low-pathogenicity vaccine strain. *J. Virol.* **79**, 12401–12407 (2005).
- Zeng, H. *et al.* Highly pathogenic avian influenza H5N1 viruses elicit an attenuated type I interferon response in polarized human bronchial epithelial cells. *J. Virol.* **81**, 12439–12449 (2007).
- Zeng, H. *et al.* Tropism and infectivity of influenza virus, including highly pathogenic avian H5N1 virus, in ferret tracheal differentiated primary epithelial cell cultures. *J. Virol.* **87**, 2597–2607 (2013).
- Van Hoeven, N. *et al.* Pathogenesis of 1918 pandemic and H5N1 influenza virus infections in a guinea pig model: antiviral potential of exogenous alpha interferon to reduce virus shedding. *J. Virol.* **83**, 2851–2861 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the China CDC as part of the WHO Global Influenza Surveillance and Response System (GISRS) for facilitating access to viruses and A. Balish for preparation of viruses. The findings and conclusions in this report are those of the authors and do not necessarily reflect the views of the funding agency. Glycan microarray slides were produced under contract for the CDC using a glycan library generously provided by the Consortium for Functional Glycomics, funded by National Institute of General Medical Sciences Grant GM62116.

Author Contributions J.A.B., K.M.G., T.R.M., H.Z., X.S., P.J.C., J.S., J.M.K., and T.M.T. designed the experiments; J.A.B., K.M.G., M.B.P., T.R.M., H.Z., C.P., X.S., P.J.C., J.S., and T.M.T. performed the experiments; J.A.B., K.M.G., T.R.M., H.Z., C.P., X.S., P.J.C., J.M.V., J.S., J.M.K., and T.M.T. analysed data; J.A.B., T.R.M., H.Z., C.P., P.J.C., J.S., and J.M.K., T.M.T. wrote the manuscript.

Author Information GISAID (<http://platform.gisaid.org/>) accession numbers are as follows for ck/VN/11 (PB2, EPI457480; PB1, EPI457481; PA, EPI457479; HA, EPI457483; NP, EPI457476; NA, EPI457482; MP, EPI457478; NS, EPI457477) and shv/Egypt/07 (PB2, EPI372407; PB1, EPI372408; PA, EPI372406; HA, EPI372410; NP, EPI372403; NA, EPI372409; MP, EPI372405; NS, EPI372404). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.M.T. (tft9@cdc.gov).

METHODS

Viruses. All animal experiments were conducted under biosafety level 3 facilities with enhancements (BSL3+) in accordance with US Federal and World Health Organization guidelines. Biocontainment for the BSL3+ laboratories is based on the 5th Edition of Biosafety in Microbiological and Biomedical Laboratories. A/Anhui/1/13 (H7N9), A/Shanghai/1/13 (H7N9), A/shoveler/Egypt/00215-NAMRU3/07 (H7N9), A/chicken/Vietnam/NCVD-1156/2011 (H9N2) and A/Goose/Nebraska/17097-4/11 (H7N9) viruses were propagated in 10-day-old embryonated hens' eggs for 42 h at 37 °C as previously described^{18,23}. A/Texas/50/12 was propagated in Madin Darby canine kidney (MDCK) cells as previously described²⁴.

Ferret experiments. All animal research was approved by the Centers for Disease Control and Prevention's Institutional Animal Care and Use Committee and conducted in an Association for Assessment and Accreditation of Laboratory Animal Care International-accredited animal facility. Male Fitch ferrets (Triple F Farms), 6 to 7 months of age and serologically negative by haemagglutination inhibition (HI) assay for currently circulating influenza viruses, were used in all experiments. A minimum number of animals were used to achieve reproducible results in each experiment per ethical guidelines; investigators were not blinded to groups. Ferrets were housed in cages within a Duo-Flow Bioclean mobile clean room (Lab Products) and randomly assigned to experimental groups. Temperatures were measured using a subcutaneous implantable temperature transponder (BioMedic Data Systems). Ferrets were inoculated intranasally with 10⁶ p.f.u. of virus in a 1 ml volume diluted in PBS, then monitored daily for changes in body temperature, weight and clinical signs of infection. Lethargy was measured as described previously¹⁴. Any ferret that lost >25% of its pre-inoculation body weight or showed neurological dysfunction was euthanized and submitted to post-mortem examination. Nasal washes, conjunctival washes and rectal swabs were collected on alternate days post-inoculation (p.i.) or post-contact, immediately frozen on dry ice and stored at -70 °C until titration as previously described²⁵. Tissue specimens collected for virus titration were immediately frozen on dry ice and stored at -70 °C until titration as previously described¹⁸. Blood was collected from inoculated ferrets on indicated days p.i. in EDTA Vacutainer tubes (BD) and complete blood counts were quantified using a Hemavet HV950FS instrument per the manufacturer's instructions (Drew Scientific).

For transmission experiments, ferrets were inoculated as described above. Approximately 24 h after inoculation, a naive ferret was placed in the same cage as an inoculated animal (for assessment of virus transmission in the presence of direct contact) or in an adjacent cage with perforated side-walls (holes 1–5 mm in diameter, with cages spaced ~3 mm apart) to allow for air exchange between cages in the absence of direct or indirect contact (for assessment of virus transmission by respiratory droplets). The term 'respiratory droplet transmission' refers to virus transmission in the absence of direct or indirect contact and does not imply an understanding of the size of droplets or aerosols involved in the transmission events. To prevent inadvertent transmission in direct contact or respiratory droplet experiments, investigators always handled the contact animals first, and decontaminated all surfaces which came into contact with each ferret before handling the next animal.

Whole-genome sequencing of nasal washes. RNA was extracted from nasal wash samples from the peak day of virus shedding presenting in the inoculated and respiratory droplet contact (2/6 for Anhui/1 ferrets, and 1/3 for Shanghai/1 virus experiment) ferrets and subjected to RT-PCR (PCR with reverse transcription) followed by DNA sequencing analysis. Obtained sequences obtained from all virus gene segments were compared with relevant virus sequences available from The Global Initiative on Sharing All Influenza Data (GISAID) database.

Mouse experiments. Female BALB/c mice 6 to 9 weeks old were anaesthetized with an intraperitoneal injection of 0.15 ml of 2,2,2-tribromoethanol in *tert*-amyl alcohol (Avertin; Sigma-Aldrich) prior to virus inoculation. Intranasal inoculations were performed by instilling 50 µl of virus diluted in PBS onto the nares of the animals. The 50% mouse infectious dose (MID₅₀, determined by virus detection in

the lungs day 3 p.i.) and 50% lethal dose (LD₅₀) were determined by inoculating groups of eight mice with serial 10-fold dilutions of virus. Three mice per dilution were euthanized on day 3 p.i.; lungs were collected from each mouse, immediately frozen on dry ice, and stored at -70 °C. Tissues were thawed, homogenized in 1 ml cold PBS, and clarified by centrifugation prior to titration by standard plaque assay. Five mice per dilution were observed daily for 14 days for morbidity (as measured by weight loss) and mortality; any mouse which lost >25% of its pre-inoculation weight was euthanized. MID₅₀ and LD₅₀ values were determined following ref. 26. To investigate the ability of each virus to spread systemically in mice, an additional three mice were euthanized at days 3 and 6 p.i. from mice inoculated with 10⁶ p.f.u. of each virus for collection of nose, lung and brain tissue, which were stored and processed as described above.

Ocular inoculations were performed by lightly scarifying the right eye of the mouse with a corneal trephine and instilling 5 µl of virus (10⁶ p.f.u.) onto the surface of the eye and massaging the virus into the eye with the eyelid as previously described²⁷. An additional three mice were inoculated and euthanized on days 3 and 6 p.i. for collection of eye, nose and lung tissue, which were stored and processed as described above.

Cell culture and viral replication. The human bronchial epithelial cell line Calu-3 (ATCC) was grown on membrane inserts as previously described²⁰. Cells were grown to confluence in six-well plates for 1 week until the confluent monolayer reached a stable transepithelial resistance of >1,000 Ω cm². Virus was added to cells apically in serum-free medium at a multiplicity of infection (MOI) of 0.01 for one hour before washing. Plates were incubated at either 37 °C or 33 °C for the duration of the experiment as indicated. Aliquots of culture supernatant taken p.i. were immediately frozen at -80 °C until titration. All samples collected for replication kinetics were titrated for the presence of infectious virus by standard plaque assay in MDCK cells.

Serology. HI assays were performed with ferret sera collected 16–20 days after inoculation or exposure to infected animals using 0.5% turkey erythrocytes against homologous virus²⁸.

Cytokine and chemokine quantification. Clarified mouse lung tissue homogenates were analysed with the BioPlex protein array system (Bio-Rad), according to manufacturer's instructions. A minimum of three independent samples were collected and tested in duplicate for each condition.

Glycan microarray analysis. Microarray printing and influenza virus analyses have been described previously¹¹. A/Anhui/1/2013 virus was analysed at a HA titre of 512 while A/Shanghai/1/2013 was analysed at an HA titre of 256. Specific details on the glycans analysed in these studies are provided in Supplementary Table 1.

Statistics. Statistical significance ($P < 0.05$) for murine and ferret studies was determined by one-way analysis of variance (ANOVA) with a Bonferroni post-test. Statistical significance for *in vitro* studies was determined using the Mann-Whitney and Student's *t* tests. All *in vitro* studies were conducted with sufficient sample size for statistical analyses performed; indicated tests were chosen as appropriate for each data set to be analysed and all data meet the assumptions of the tests.

23. Belser, J. A. *et al.* Pathogenesis, transmissibility, and ocular tropism of a highly pathogenic avian influenza A (H7N3) virus associated with human conjunctivitis. *J. Virol.* **87**, 5746–5754 (2013).
24. Tumpey, T. M. *et al.* Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science* **310**, 77–80 (2005).
25. Belser, J. A. *et al.* Influenza virus respiratory infection and transmission following ocular inoculation in ferrets. *PLoS Pathog.* **8**, e1002569 (2012).
26. Reed, L. J. & Muench, H. A simple method of estimating fifty per cent endpoints. *Am. J. Hyg.* **27**, 493–497 (1938).
27. Belser, J. A., Wadford, D. A., Xu, J., Katz, J. M. & Tumpey, T. M. Ocular infection of mice with influenza A (H7) viruses: a site of primary replication and spread to the respiratory tract. *J. Virol.* **83**, 7075–7084 (2009).
28. Stephenson, I., Wood, J. M., Nicholson, K. G., Charlett, A. & Zambon, M. C. Detection of anti-H5 responses in human sera by HI using horse erythrocytes following MF59-adjuvanted influenza A/Duck/Singapore/97 vaccine. *Virus Res.* **103**, 91–95 (2004).

Limited airborne transmission of H7N9 influenza A virus between ferrets

Mathilde Richard^{1*}, Eefje J. A. Schrauwen^{1*}, Miranda de Graaf¹, Theo M. Bestebroer¹, Monique I. J. Spronken¹, Sander van Boheemen¹, Dennis de Meulder¹, Pascal Lexmond¹, Martin Linster¹, Sander Herfst¹, Derek J. Smith², Judith M. van den Brand¹, David F. Burke², Thijs Kuiken¹, Guus F. Rimmelzwaan¹, Albert D. M. E. Osterhaus¹ & Ron A. M. Fouchier¹

Wild waterfowl form the main reservoir of influenza A viruses, from which transmission occurs directly or indirectly to various secondary hosts, including humans¹. Direct avian-to-human transmission has been observed for viruses of subtypes A(H5N1), A(H7N2), A(H7N3), A(H7N7), A(H9N2) and A(H10N7) upon human exposure to poultry^{2–7}, but a lack of sustained human-to-human transmission has prevented these viruses from causing new pandemics. Recently, avian A(H7N9) viruses were transmitted to humans, causing severe respiratory disease and deaths in China⁸. Because transmission via respiratory droplets and aerosols (hereafter referred to as airborne transmission) is the main route for efficient transmission between humans, it is important to gain an insight into airborne transmission of the A(H7N9) virus. Here we show that although the A/Anhui/1/2013 A(H7N9) virus harbours determinants associated with human adaptation and transmissibility between mammals, its airborne transmissibility in ferrets is limited, and it is intermediate between that of typical human and avian influenza viruses. Multiple A(H7N9) virus genetic variants were transmitted. Upon ferret passage, variants with higher avian receptor binding, higher pH of fusion, and lower thermostability were selected, potentially resulting in reduced transmissibility. This A(H7N9) virus outbreak highlights the need for increased understanding of the determinants of efficient airborne transmission of avian influenza viruses between mammals.

At the end of March 2013, the World Health Organization was notified by the Chinese authorities of three human cases of infection with a novel avian-origin influenza A(H7N9) virus⁹. All three cases developed bilateral pneumonia with progression to acute respiratory distress syndrome and death^{10,11}. As of July 2013, 132 A(H7N9) laboratory-confirmed cases have been reported in ten different provinces of China, including 37 deaths⁸. This novel A(H7N9) virus emerged in humans after reassortment between viruses of poultry and wild bird origin. The haemagglutinin (HA) and neuraminidase (NA) genes are genetically related to H7 and N9 of viruses isolated from wild ducks, whereas the other genes are closely related to A(H9N2) viruses circulating in poultry¹². It is most likely that the new A(H7N9) viruses have circulated undetected in domestic birds because of their low pathogenicity for poultry. A(H7N9) viruses were isolated from specimens at live poultry markets, pointing to domestic birds as a potential source of human infections¹³.

Although A(H7N9) viruses harbour genetic traits associated with human adaptation of avian viruses and increased transmission between mammals, such as the Q217L substitution in HA (H7 numbering, 226 in H3 numbering) conferring a human receptor preference¹⁴ and the E627K substitution in PB2¹⁵, no sustained human-to-human transmission of A(H7N9) viruses has been reported so far. Apart from two confirmed cases that might have arisen from family clusters and for which human-to-human transmission cannot be ruled out, human cases of A(H7N9) infection were epidemiologically unrelated and identified in different parts of China¹¹.

Gaining knowledge about the ability of animal viruses to transmit via the airborne route is crucial to be able to mitigate future pandemics. Recently, the airborne transmissibility of the human isolate A/Shanghai/2/2013 was evaluated in ferrets and in pigs and was found to be less robust than for the 2009 pandemic A(H1N1) (pH1N1) virus¹⁶. Here, we assessed the airborne transmissibility of a different human isolate, A/Anhui/1/2013 (Anhui/1), in the ferret model as described^{17,18}. Both A/Shanghai/2/2013 and A/Anhui/1/2013 possess human receptor specificity and are of particular interest regarding transmission. Four donor ferrets were inoculated intranasally with 10⁶ 50% tissue culture infectious doses (TCID₅₀) of Anhui/1 virus isolate and the following day, four recipient ferrets were placed in adjacent cages, designed to prevent direct contact between animals but to allow transmission via the airborne route. Throat and nasal swabs were collected at 1, 3, 5 and 7 days post-inoculation (d.p.i.) from the donor ferrets and at 1, 3, 5, 7 and 9 days post-exposure (d.p.e.) from the recipient ferrets. Virus shedding from the donor ferrets was detected from 1 d.p.i. onwards with infectious virus titres up to 10^{6.5} TCID₅₀ ml⁻¹ (Fig. 1a–d). Anhui/1 virus was transmitted to three out of four recipient ferrets (F1, F2 and F3). Transmission was detected at 3 d.p.e. for two ferrets (F2 and F3) and at 5 d.p.e. for one ferret (F1) with infectious virus titres in respiratory swabs up to 10⁶ TCID₅₀ ml⁻¹. All three animals infected via the airborne route seroconverted by 2 weeks after exposure, whereas recipient ferret F4 did not seroconvert. Using the same experimental set-up and protocol, we previously tested the transmissibility of the human pH1N1 virus, seasonal human A(H1N1) virus, and avian influenza A(H5N1) virus. Whereas A(H5N1) virus was not transmitted between ferrets via the airborne route, human influenza viruses were transmitted in all donor–recipients pairs^{17,18}. Replication in donor ferrets inoculated with the Anhui/1 virus and pandemic and seasonal A(H1N1) viruses was comparable, but virus shedding from recipient ferrets was less abundant and delayed for the Anhui/1 virus as compared to the A(H1N1) viruses.

Sanger sequencing was used to determine the consensus genome sequence of the three airborne-transmitted Anhui/1 viruses isolated from recipient ferrets F1, F2 and F3 (Table 1), and substitutions were detected in all gene segments. Several of these were already present in the inoculum, demonstrating the presence of a mixture of viruses in the Anhui/1 virus isolate. Consultation with other laboratories revealed that heterogeneous virus mixtures were present in A(H7N9) virus isolates shipped to other laboratories also. Viruses with different genotypes were transmitted (Table 1). Only two substitutions, N123D and N149D in HA, were consistently found in all three airborne-transmitted viruses. These substitutions are not part of potential N-linked glycosylation sites. Virus recovered from the recipient ferret F1 at 7 d.p.e., which contained the lowest number of substitutions compared to the Anhui/1 virus isolate and had a high virus load, was used to inoculate four additional ferrets. One day later, these animals were paired with four recipient ferrets in transmission cages (Fig. 1e–h), of which one became infected (F5).

¹Department of Viroscience, Erasmus Medical Center, 3015GE Rotterdam, The Netherlands. ²Department of Zoology, University of Cambridge, CB2 3EJ Cambridge, UK.

*These authors contributed equally to this work.

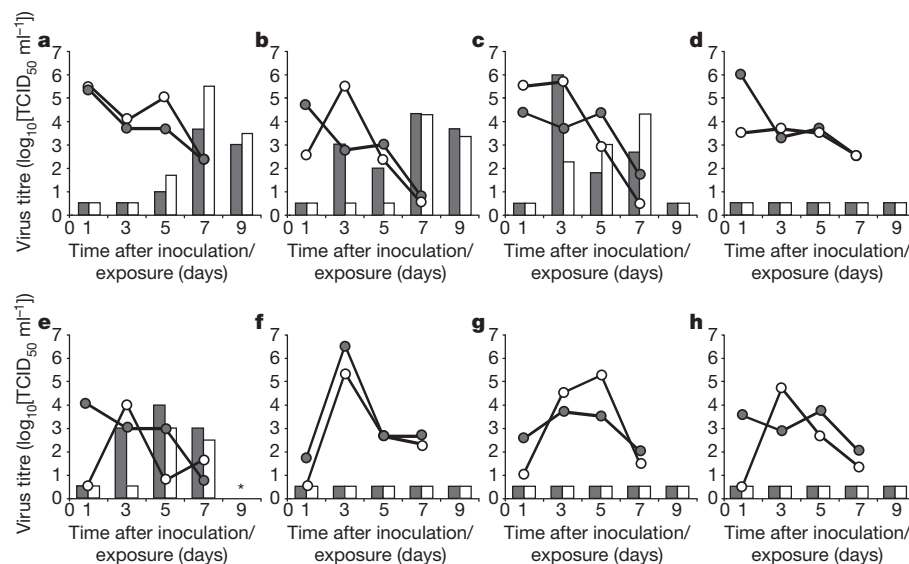


Figure 1 | Airborne transmission of Anhui/1 viruses between ferrets. **a–d**, Transmission experiments are shown for Anhui/1 virus isolate in four ferret pairs (F1–F4). **e–h**, A nose swab sample from the recipient ferret F1 at 7 d.p.e. was used for the transmission experiments in four ferret pairs shown (F5–F8). Data for individual transmission experiments are shown in each panel, with virus shedding in inoculated and airborne virus-exposed animals shown as lines and bars, respectively. Black circles and bars represent shedding from the throat; white circles and bars represent shedding from the nose. The asterisk indicates the lack of swab collection at 9 d.p.e. for the recipient ferret euthanized at 8 d.p.e. The lower limit of detection is $0.5 \log_{10} \text{TCID}_{50} \text{ ml}^{-1}$.

None of the donor ferrets developed clinical signs upon intranasal inoculation. Two recipient ferrets were also without clinical signs, but one recipient ferret (F2) showed loss of appetite, ruffled fur, lethargy and breathing difficulties from 7 d.p.e. onwards. Recipient ferret F5 became moribund and was euthanized at 8 d.p.e. Infectious virus titres up to $10^{4.9} \text{TCID}_{50} \text{ g}^{-1}$ in the lungs and $10^{5.8} \text{TCID}_{50} \text{ g}^{-1}$ in nasal turbinates were detected in ferret F5. On the basis of pathological analysis, this animal was suffering from a moderate multifocal suppurative rhinitis with associated virus antigen expression in the nasal respiratory and olfactory epithelium (Supplementary Table 4). No lesions or virus antigen expression were seen in the other parts of the respiratory tract, liver or brain that could explain the lethargic state of the ferret.

Consensus virus genome sequences as determined by Sanger sequencing of the airborne-transmitted viruses recovered from recipient ferrets F1 and F5 were identical (Table 1). Substitutions N123D and N149D in HA and M523I in the basic polymerase 1 (PB1) were found consistently in two subsequent transmission experiments. Because mixed populations were detected in the inoculum using Sanger sequencing, next-generation sequencing was performed with respiratory samples of the donor–recipient pairs for which airborne transmission was observed. The entire HA gene and the PB1 part that contained the M523I substitution (nucleotide positions 1126–1616) were sequenced using a 454

sequencing platform (Roche). None of the airborne-transmitted viruses possessed consensus genome sequences identical to that of the Anhui/1 virus isolate (Table 2). Interestingly, an L217Q substitution, conferring a receptor switch from human to avian specificity ($\alpha 2.6$ - to $\alpha 2.3$ -linked sialic acids, respectively), was detected in donor ferrets F1, F2 and F3 in 9.1%, 15.7% and 24.5% of the total number of reads, respectively, but this substitution was not detected in any of the viruses isolated from the ferrets upon airborne exposure. By setting the detection threshold at 1%, we confirmed that the number of variable nucleotide positions in the genome of the Anhui/1 virus isolate was high, and that a rapid gain in clonality occurred after two transmission experiments (Table 2 and Supplementary Table 1). The double substitution N123D/N149D in HA appeared to be selected in most airborne-transmitted viruses and constituted the main viral population after two subsequent transmission experiments (Supplementary Table 2). Although there appeared to be selection of N123D and N149D, it was not possible to determine whether the selection occurred at the level of individual or double substitutions. However, the rapid selection of substitutions in the HA and PB1 genes and the gain in clonality did not change transmission substantially enough to be detectable with the current group size of four ferrets.

Residues N123 and N149 are adjacent to the receptor binding site but do not interact directly with $\alpha 2.3$ - and $\alpha 2.6$ -linked sialic acids

Table 1 | Sanger sequence analysis of full viral genomes of the Anhui/1 virus inoculum and airborne-transmitted viruses

Segment	nt position	nt WT	nt mut.	Amino acid position	Amino acid WT	Amino acid mut.	Inoculum	Transmission 1			Transmission 2	
								Recipients				
								F1	F2	F3	F5	
PB2	411	A	G	128	Gly	S					X	
PB2	1017	C	T	330	Phe	S	X*				X	
PB2	1309	C	T	428	Leu	S		X				X
PB2	1846	C	T	607	Leu	S	X*				X	
PB1	1593	G	A	523	Met	Ile		X				X
PB1	2055	G	A	678	Ser	Asn				X		
PA	1070	A	G	349	Glu	Gly				X		
PA	1167	C	T	380	Asp	S				X		
PA	1380	C	T	452	His	S					X	
PA	1616	G	A	531	Arg	Lys					X	
PA	1674	G	T	550	Leu	S				X*		
PA	1776	C	T	584	Cys	S				X*		
HA	442	A	G	123	Asn	Asp	X*	X	X*	X		X
HA	520	A	G	149	Asn	Asp	X*	X	X*	X		X
HA	704	C	A	210	Ala	Glu				X*		
NP	718	A	C	225	Ile	Leu				X*		
NA	46	C	T	10	Thr	Ile	X*			X*	X*	
M	652	T	C	225	Ala	S				X*		
NS	180	C	T	52	Leu	S					X*	

Substitutions in bold were found in two subsequent transmission experiments and were phenotypically characterized. nt, nucleotides; S, silent substitution; WT, wild type.

* Mixture of wild-type and mutant nucleotides.

Table 2 | Amino acid substitutions in the HA gene and the PB1 gene in Anhui/1 viruses before and after transmission in ferrets

Gene	nt position	nt WT	nt mut.	Amino acid position	Amino acid WT	Amino acid mut.	Inoculum* (%/no. reads)	Donor* F1 (%/no. reads)	Recipient* F1 (%/no. reads)	Donor* F2 (%/no. reads)	Recipient* F2 (%/no. reads)	Donor* F3 (%/no. reads)	Recipient* F3 (%/no. reads)	Donor* F5 (%/no. reads)	Recipient* F5 (%/no. reads)
HA	117	C	T	14	Thr	S	19.5/6,629	–	–	12.5/7,954	20.3/10,349	7.2/4,763	9.6/5,364	–	–
HA	257	C	T	61	Thr	Ile	5.6/8,142	–	–	–	–	–	–	–	–
HA†	442†	A†	G†	123†	Asn†	Asp†	48.7/8,648†	14.4/9,183†	99.8/6,224†	37.4/11,196†	28.5/12,139†	36.5/6,810†	90.1/8,417†	99.8/11,576†	99.9/10,182†
HA	448	G	A	125	Ala	Thr	1.8/7,740	–	–	14.7/10,044	20.5/10,733	7.6/6,088	10.4/7,250	–	–
HA†	520†	A†	G†	149†	Asn†	Asp†	62.5/6,163†	17.9/6,737†	99.9/3,622†	50.4/8,447†	28.6/7,818†	61.2/5,266†	90.4/6,840†	99.7/9,229†	99.7/9,047†
HA	704	C	A	210	Ala	Glu	–	72.5/3,903	–	28.4/4,804	56.4/4,109	8.5/3,359	–	–	–
HA‡	725‡	T‡	A‡	217‡	Leu‡	Gln‡	–	9.1/3,831‡	–	15.7/4,749‡	–	24.5/3,322‡	–	–	–
HA	1032	G	T	319	Lys	Asn	–	–	–	8.3/6,447	–	–	–	–	–
HA	1218	C	T	381	Asn	S	–	5.2/2,102	–	–	–	23.4/1,963	–	–	–
HA	1396	G	A	441	Glu	Lys	–	–	–	–	14.5/2,299	–	–	–	–
HA	1422	G	A	449	Glu	S	–	–	–	–	–	12.6/4,027	–	–	–
HA	1575	C	T	500	Ser	S	6.1/6,532	–	–	–	15.5/2,245	–	7.8/2,387	–	–
HA	1706	T	C	NCR	–	–	–	–	93.0/470	–	–	–	–	–	–
PB1§	1404	G	A	460	Gln	S	6.8/603	–	–	–	–	11.9/1,024	–	–	–
PB1†	1593†	G†	A†	523†	Met†	Ile†	–	–	100/256†	–	–	–	–	82.2/1,795†	99.9/1,064†

NCR, non-coding region; S, silent substitution.

*Percentage variant present with a detection threshold of 5% and number of reads is shown.

†Substitutions were found in two subsequent transmission experiments and were phenotypically characterized.

‡Substitution (L217Q) corresponds to the receptor switch from α 2,6- to α 2,3-linked sialic acids preference.

§The PB1 gene was partially sequenced (nucleotide positions 1126–1616).

(Fig. 2). Recombinant viruses were generated based on seven gene segments of A/Puerto Rico/8/1934 (Puerto Rico/8) with the wild-type Anhui/1 HA and the Anhui/1 HA with amino acid substitutions N123D (Anhui/1(N123D)), N149D (Anhui/1(N149D)) and N123D plus N149D (Anhui/1(N123D/N149D)). Binding to α 2,3-linked sialic acids of Anhui/1(N123D), Anhui/1(N149D) and Anhui/1(N123D/N149D) viruses as assessed using a modified turkey red blood cell (TRBC) assay¹⁹ was increased slightly by two- to fourfold compared to Anhui/1 virus (Supplementary Table 3).

It was recently proposed that stability of HA in an acidic environment—such as mammalian nasal mucosa—is a determinant for airborne transmissibility of influenza viruses²⁰. It has also been noted that

viruses that fuse at low pH have higher thermostability than those fusing at higher pH²¹. We observed that fusion for Anhui/1 HA occurred at pH 5.6, a higher pH than reported previously for human viruses²². Neither the single N123D and N149D substitutions nor the double N123D/N149D substitution reduced the threshold pH for HA-mediated membrane fusion compared to the wild-type Anhui/1 HA (Supplementary Fig. 1). Both single substitutions N123D and N149D and the double substitution N123D/N149D decreased the temperature stability compared to Anhui/1 (Supplementary Fig. 2).

We also assessed the effect of the M523I substitution on the polymerase complex activity in mammalian cells using a minigenome assay at 33 °C, the temperature of the mammalian upper respiratory tract,

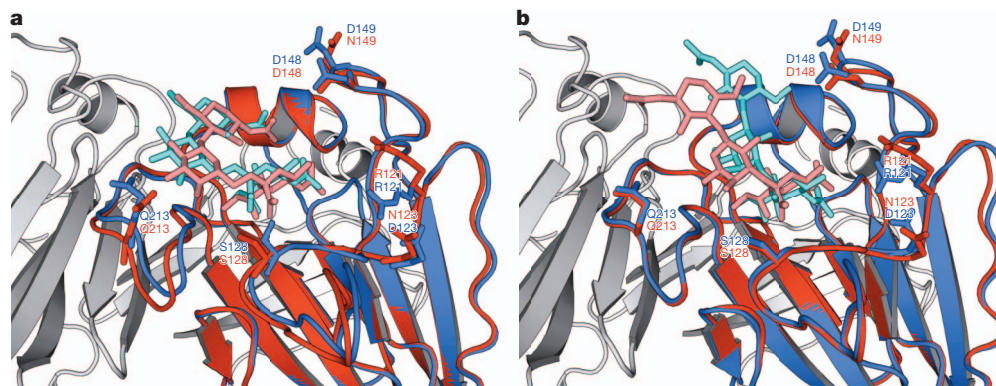


Figure 2 | Cartoon representation of a model of the trimer structure of HA of Anhui/1 (red) and Anhui/1(N123D,N149D) (blue) bound to α 2,6- (a) and α 2,3-linked sialic acids (b). The structures of the three-sugar glycans NeuAc α 2,6Gal β 1-4GlcNAc (a) and NeuAc α 2,3Gal β 1-4GlcNAc (b) were docked into the H7 receptor binding site. The glycans and the amino acid substitutions discussed in the text are shown as sticks. Amino acids N123 and N149 are adjacent to the receptor binding site and, in Anhui/1, do not interact directly with the three-sugar glycans that are depicted in the figure. The mutations cause small changes in the position of some of the residues around

the receptor binding site, notably R121 and D148, and additionally for the α 2,6-linked glycan, residues S128 and Q213. The D123 mutant can form stronger interactions with the side chain of R121, restricting the movement and orientating its side chain to point towards the receptor binding site and interact with the glycan. In Anhui/1, N149 interacts with the neighbouring residue, D148, restricting its orientation. The D149 mutant allows the side chain of D148 to rearrange and interact with the glycan. These changes allow both the α 2,6- and α 2,3-linked glycans to alter position and form more interactions with the HA. All residues are labelled in H7 numbering.

and at 37 °C as previously described²³. No differences in polymerase activity were observed for the polymerase complex with or without the M523I substitution in PB1 (Supplementary Fig. 3).

Here we report that airborne transmission of Anhui/1, as for A/Shanghai/2/2013, can occur between ferrets. Keeping in mind that quantifying transmission in our current experimental model is difficult, and that any transmission may not be directly extrapolated to transmission between humans, these data suggest that Anhui/1 transmission is more efficient than for other avian influenza viruses, which are not airborne transmitted in ferrets, but less robust—with fewer animals becoming infected, and less and delayed virus shedding—as compared to seasonal and pandemic A(H1N1) virus transmission¹⁷. Despite efficient virus replication in ferrets (Fig. 1) and attachment to α 2.6-linked sialic acids (Supplementary Table 3), we speculate that the residual binding to α 2.3-linked sialic acids (Supplementary Table 3), the fusion occurring at a relatively high pH (Supplementary Fig. 1), and instability of HA (Supplementary Fig. 2) may be responsible, at least in part, for limited transmission of Anhui/1 A(H7N9) virus. Contrary to what was observed for A(H5N1) virus^{18,20}, the substitutions selected upon ferret passage and transmission of A(H7N9) virus—N123D and N149D—increased binding to α 2.3-linked sialic acids, increased the pH threshold for membrane fusion, and decreased the thermostability of HA, thus not contributing to increased virus transmission.

Influenza viruses carrying human adaptation markers can arise in poultry. This seems to be the case for the newly emerged A(H7N9) viruses, but also during the 2003 A(H7N7) outbreak in the Netherlands²⁴ and for A(H5N1) viruses currently circulating in poultry²⁵. Fortunately, additional changes required to, for example, further tune receptor preference, lower the pH for HA fusion, and increase HA stability, may be needed for the A(H7N9) viruses to transmit efficiently in mammals^{18,20}. Increased understanding of the mechanisms and molecular determinants that facilitate crossing the species barrier and airborne transmission of avian influenza viruses between mammals is urgently needed.

METHODS SUMMARY

Viruses. Influenza virus Anhui/1 was isolated from a human case of infection and passaged three times in embryonated chicken eggs and once in Madin–Darby canine kidney (MDCK) cells. The virus was provided by the Chinese CDC via the WHO collaborating centre in the UK under the WHO PIP framework. A synthetic construct of the Anhui/1 HA gene segment was provided by R. Webby. The HA gene segment was cloned in the pCAGGs expression plasmid. The PB2, PB1, PA and NP gene segments were cloned in a modified version of the bidirectional reverse genetics plasmid pHW2000^{26,27}. Mutations of interest (M523I in PB1, D123N and D149N in HA) were introduced in constructs using the QuikChange multi-site-directed mutagenesis kit (Stratagene). Recombinant viruses containing seven gene segments of A/Puerto-Rico/8/1934 and wild-type or mutant Anhui/1 HA were produced upon transfection of 293T cells. Virus stocks were propagated and titrated in MDCK cells as described²⁶.

Ferret models. Transmission experiments were performed as described^{17,18}. In short, four female seronegative ferrets, 1–2 years of age, were inoculated intranasally with 10^6 TCID₅₀ of virus. Each donor ferret was then placed in a transmission cage. One day after inoculation, one naive recipient ferret was placed opposite each donor ferret. Each transmission pair was housed in a separate transmission cage designed to prevent direct contact but allowing airflow from donor to recipient. Nose and throat swabs were collected on 1, 3, 5 and 7 d.p.i. for donor ferrets and on 1, 3, 5, 7 and 9 d.p.e. for recipient ferrets. Virus titres in swabs were determined by end-point titration in MDCK cells. A nose swab sample of recipient ferret F1 at 7 d.p.e. was used for the second transmission experiment, with a final dose of approximately 10^4 TCID₅₀ for each ferret.

Full Methods and any associated references are available in the online version of the paper.

Received 6 June; accepted 17 July 2013.

Published online 7 August 2013.

1. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **56**, 152–179 (1992).

2. Beigel, J. H. *et al.* Avian influenza A (H5N1) infection in humans. *N. Engl. J. Med.* **353**, 1374–1385 (2005).
3. Hirst, M. *et al.* Novel avian influenza H7N3 strain outbreak, British Columbia. *Emerg. Infect. Dis.* **10**, 2192–2195 (2004).
4. Fouchier, R. A. *et al.* Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *Proc. Natl Acad. Sci. USA* **101**, 1356–1361 (2004).
5. Ostrowsky, B. *et al.* Low pathogenic avian influenza A (H7N2) virus infection in immunocompromised adult, New York, USA, 2003. *Emerg. Infect. Dis.* **18**, 1128–1131 (2012).
6. Chen, Y. *et al.* Human infections with the emerging avian influenza A H7N9 virus from wet market poultry: clinical analysis and characterisation of viral genome. *Lancet* **381**, 1916–1925 (2013).
7. Arzey, G. G. *et al.* Influenza virus A (H10N7) in chickens and poultry abattoir workers, Australia. *Emerg. Infect. Dis.* **18**, 814–816 (2012).
8. WHO. Avian Influenza A(H7N9) Virus http://www.who.int/influenza/human_animal_interface/influenza_h7n9/en/index.html (2013).
9. WHO. Human Infection with Influenza A(H7N9) Virus in China http://www.who.int/csr/don/2013_04_01/en/index.html (1 April 2013).
10. Gao, R. *et al.* Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.* **368**, 1888–1897 (2013).
11. Li, Q. *et al.* Preliminary report: Epidemiology of the avian influenza A(H7N9) outbreak in China. *N. Engl. J. Med.* <http://dx.doi.org/10.1056/NEJMoa1304617> (2013).
12. Liu, D. *et al.* Origin and diversity of novel avian influenza A H7N9 viruses causing human infection: phylogenetic, structural, and coalescent analyses. *Lancet* **381**, 1926–1932 (2013).
13. Cowling, B. J. *et al.* Comparative epidemiology of human infections with avian influenza A H7N9 and H5N1 viruses in China: a population-based study of laboratory-confirmed cases. *Lancet* **382**, 129–137 (2013).
14. Xiong, X. *et al.* Receptor binding by an H7N9 influenza virus from humans. *Nature* <http://dx.doi.org/10.1038/nature12372> (20 June 2013).
15. Steel, J., Lowen, A. C., Mubareka, S. & Palese, P. Transmission of influenza virus in a mammalian host is increased by PB2 amino acids 627K or 627E/701N. *PLoS Pathog.* **5**, e1000252 (2009).
16. Zhu, H. *et al.* Infectivity, transmission, and pathology of human H7N9 influenza in ferrets and pigs. *Science* **341**, 183–186 (2013).
17. Munster, V. J. *et al.* Pathogenesis and transmission of swine-origin 2009 A(H1N1) influenza virus in ferrets. *Science* **325**, 481–483 (2009).
18. Herfst, S. *et al.* Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* **336**, 1534–1541 (2012).
19. Chutinimitkul, S. *et al.* *In vitro* assessment of attachment pattern and replication efficiency of H5N1 influenza A viruses with altered receptor specificity. *J. Virol.* **84**, 6825–6833 (2010).
20. Imai, M. *et al.* Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* **486**, 420–428 (2012).
21. Ruigrok, R. W. *et al.* Conformational changes in the hemagglutinin of influenza virus which accompany heat-induced fusion of virus with liposomes. *Virology* **155**, 484–497 (1986).
22. Galloway, S. E., Reed, M. L., Russell, C. J. & Steinhauer, D. A. Influenza HA subtypes demonstrate divergent phenotypes for cleavage activation and pH of fusion: implications for host range and adaptation. *PLoS Pathog.* **9**, e1003151 (2013).
23. Subbarao, E. K., London, W. & Murphy, B. R. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *J. Virol.* **67**, 1761–1764 (1993).
24. Jonges, M. *et al.* Comparative analysis of avian influenza virus diversity in poultry and humans during a highly pathogenic avian influenza A (H7N7) virus outbreak. *J. Virol.* **85**, 10598–10604 (2011).
25. Russell, C. A. *et al.* The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. *Science* **336**, 1541–1547 (2012).
26. de Wit, E. *et al.* Efficient generation and growth of influenza virus A/PR/8/34 from eight cDNA fragments. *Virus Res.* **103**, 155–161 (2004).
27. Hoffmann, E., Neumann, G., Kawaoka, Y., Hobom, G. & Webster, R. G. A DNA transfection system for generation of influenza A virus from eight plasmids. *Proc. Natl Acad. Sci. USA* **97**, 6108–6113 (2000).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. van Run, S. van der Vliet and A. Reiners for technical assistance. We thank the Chinese CDC for providing the A/Anhui/1/2013 isolate and R. Webby for the synthetic construct of HA. This work was financed through NIAID-NIH contract HHSN266200700010C and EU FP7 programs EMPIRE and ANTIGONE.

Author Contributions M.R., E.J.A.S. and R.A.M.F. designed the experiments. M.R., E.J.A.S., M.G., M.I.J.S., T.M.B., S.v.B., D.M., P.L., M.L. and S.H. performed the experiments. M.R., E.J.A.S., T.M.B., S.B., J.M.v.d.B. and D.F.B. analysed the data, and M.R., E.J.A.S., D.J.S., T.K., G.F.R., A.D.M.E.O. and R.A.M.F. wrote the manuscript. M.R. and E.J.A.S. contributed equally to the work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.A.M.F. (r.fouchier@erasmusmc.nl).

METHODS

Biocontainment. All experiments were conducted within the enhanced animal biosafety level 3 (ABSL3+) facility of Erasmus MC. The ABSL3+ facility consists of a negative pressurized (-30 Pa) laboratory in which all *in vivo* and *in vitro* experimental work is carried out in class 3 isolators or class 3 biosafety cabinets, which are also negative pressurized (<-200 Pa). Although the laboratory is considered 'clean' because all experiments are conducted in closed class 3 cabinets and isolators, special personal protective equipment, including laboratory suits, gloves and FFP3 facemasks, is used. Air released from the class 3 units is filtered by high efficiency particulate air (HEPA) filters and then leaves via the facility ventilation system, again via HEPA filters. Only authorized personnel that have received the appropriate training can access the ABSL3+ facility. For animal handling in the facilities, personnel always work in pairs. The facility is secured by procedures recognized as appropriate by the institutional biosafety officers and facility management at Erasmus MC and Dutch and United States government inspectors. Antiviral drugs (oseltamivir and zanamivir) are directly available¹⁸.

Viruses. Influenza virus A/Anhui/1/2013 (Anhui/1) was isolated from a human case of infection and passaged three times in embryonated chicken eggs and once in Madin-Darby Canine Kidney (MDCK) cells. The virus was provided by the Chinese CDC via the WHO collaborating centre in the UK in the context of the WHO PIP framework. A synthetic construct of the Anhui/1 HA gene segment was provided by R. Webby. The PB2, PB1, PA and NP gene segments were amplified by reverse transcription polymerase chain reaction (RT-PCR) from the Anhui/1 virus isolate and cloned in a modified version of the bidirectional reverse genetics plasmid pHW2000^{26,27}. In addition, the HA gene segment was cloned in pCAGGs expression plasmid. Mutations of interest (M523I in PB1, D123N and D149N in HA) were introduced in reverse genetics and pCAGGs vectors using the QuikChange multi-site-directed mutagenesis kit (Stratagene) according to the instructions of the manufacturer. Recombinant viruses containing seven gene segments of A/Puerto Rico/8/1934 and wild-type Anhui/1 HA or Anhui/1 HA containing the mutations N123D and N149D were produced upon transfection of 293T cells. Virus stocks were propagated and titrated in MDCK cells as described previously²⁶.

Cells. MDCK cells were cultured in Eagle's minimal essential medium (EMEM, Lonza) supplemented with 10% fetal bovine serum (FBS), 100 U ml⁻¹ penicillin (Lonza), 100 U ml⁻¹ streptomycin (Lonza), 2 mM glutamine (Lonza), 1.5 mg ml⁻¹ sodium bicarbonate (Lonza), 10 mM HEPES (Lonza), and non-essential amino acids (MP Biomedicals Europe). 293T cells were cultured in Dulbecco modified Eagle's medium (DMEM, Lonza) supplemented with 10% FBS, 100 U ml⁻¹ penicillin, 100 U ml⁻¹ streptomycin, 2 mM glutamine, 1 mM sodium pyruvate (Gibco), and non-essential amino acids. Vero cells were cultured in Iscove's modified Dulbecco's medium plus L-glutamine (IMDM, Lonza) supplemented with 10% FBS, 100 IU ml⁻¹ penicillin, 100 mg ml⁻¹ streptomycin and 2 mM glutamine.

Virus titration in MDCK cells. Virus titrations were performed as described previously¹⁷. Briefly, MDCK cells were inoculated with tenfold serial dilutions of virus stocks, nose swabs, throat swabs and homogenized tissue samples. Cells were washed with PBS 1 h after inoculation and cultured in 200 μ l of infection media, consisting of EMEM supplemented with 100 U ml⁻¹ penicillin, 100 U ml⁻¹ streptomycin, 2 mM glutamine, 1.5 mg ml⁻¹ sodium bicarbonate, 10 mM HEPES, non-essential amino acids, and 20 μ g ml⁻¹ trypsin (Lonza). Three days after inoculation, supernatants of cell cultures were tested for haemagglutinating activity using turkey erythrocytes as an indicator of virus replication in the cells. Infectious virus titres were calculated from four replicates each of the homogenized tissue samples, nose swabs, and throat swabs and for ten replicates of the virus stocks by the method of ref. 28.

Ferret models. An independent animal experimentation ethical review committee approved all animal studies. The Animal Experiments Committee (Dier Experimenten Commissie, DEC) judges ethical aspects of projects in which animals are involved. Research projects or educational projects involving laboratory animals can only be executed if they are approved by the DEC. The DEC considers the application and pays careful attention to the effects of the intervention on the animal, its discomfort, and weighs this against the social and scientific benefit to humans or animals. The researcher is required to keep the effects of the intervention to a minimum, based on the following: refinement, replacement, reduction. All experiments with ferrets were performed under animal biosafety level 3+ conditions in class 3 isolator cages. No method of randomization was used to determine how animals were allocated to the experimental groups. The investigator was not blinded to the group allocation during the experiments or when assessing the outcome.

Airborne transmission experiments were performed as described previously^{17,18}. In short, four female influenza virus and Aleutian disease virus seronegative adult ferrets, 1–2 years of age, were inoculated intranasally with 10⁶ TCID₅₀ of virus by applying 250 μ l of virus suspension to each nostril. The sample size of four is based

on earlier calculations for this type of experiment²⁹. Each donor ferret was then placed in a transmission cage. One day after inoculation, one naive recipient ferret was placed opposite each donor ferret. Each transmission pair was housed in a separate transmission cage designed to prevent direct contact but allowing airflow from the donor to the recipient ferret. Nose and throat swabs were collected on 1, 3, 5 and 7 d.p.i. for donor ferrets and on 1, 3, 5, 7 and 9 d.p.e. for the recipient ferrets. Virus titres in swabs were determined by end-point titration in MDCK cells. A nose swab sample of recipient ferret F1 at 7 d.p.e. was used for the second transmission experiment, with a final dose of approximately 10⁴ TCID₅₀ for each ferret. All animals were monitored daily for clinical signs. Necropsy was performed on one ferret that was moribund and due to ethical reasons was removed before the end of the experiment. Nasal turbinates, trachea, lungs, brain and liver were collected, homogenized in 3 ml of virus transport medium, after which the supernatant was collected and stored at -80°C . Virus titres in the supernatant were determined by end-point titration in MDCK cells. Duplicate samples of these tissues were fixed in 10% neutral-buffered formalin for pathological analysis.

Serology. The exposure of recipient ferrets to Anhui/1 viruses was confirmed by an haemagglutination inhibition assay using standard procedures³⁰. Briefly, blood of the recipient ferrets was collected 12–14 d.p.e. Antisera were pre-treated overnight with receptor-destroying enzyme (*Vibrio cholerae* neuraminidase) at 37 $^{\circ}\text{C}$ and incubated at 56 $^{\circ}\text{C}$ for 1 h the next day. Twofold serial dilutions of the antisera, starting at a 1:20 dilution, were mixed with 25 μ l of a virus stock containing 4 haemagglutinating units and were incubated at 37 $^{\circ}\text{C}$ for 30 min. Subsequently, 25 μ l 1% turkey erythrocytes were added and the mixture was incubated at 4 $^{\circ}\text{C}$ for 1 h. Haemagglutination inhibition was expressed as the reciprocal value of the highest dilution of the serum that completely inhibited agglutination of virus and erythrocytes.

Sequencing. Viral RNA was extracted from respiratory swab samples collected from the ferrets that were infected via the airborne route and the virus inoculum, using the High Pure RNA Isolation kit (Roche). All eight gene segments of the influenza viruses were amplified by RT-PCR using eight primer sets that cover the full viral genome which specifically amplify each gene segment³¹, and sequenced using a BigDye Terminator v3.1 Cycle sequencing kit (Applied Biosystems) and a 3130XL genetic analyser (Applied Biosystems), according to the instructions of the manufacturer. The consensus sequence was determined for viruses isolated from the following samples: virus inoculum obtained after three egg passages and one MDCK passage; recipient F1, nose swab 7 d.p.e.; recipient F2, throat swab 5 d.p.e.; recipient F3, nose swab 5 d.p.e.; and recipient F5, nose swab 5 d.p.e. Primer sequences are available upon request. Sequences were compared to reference sequences obtained from the GISAID EpiFluTM database (accession numbers EPI439503 through to EPI439510).

Viral RNA was extracted from the virus inoculum and respiratory swabs of ferrets using the High Pure RNA Isolation kit (Roche). RNA was subjected to RT-PCR, using 5 primer sets (for HA: set 1 5'-AGCAAAAGCAGGGGATACAA-3' and 5'-GTATGACTTAGTCATCTGCGG-3'; set 2 5'-GGCGGAATTGACAAGGAAGC-3' and 5'-CCACTATGATAGCAATCTCCTTCAC-3'; set 3 5'-GTGACTTTCAGTTTCAATGGGGC-3' and 5'-GATTCTCCATTGCTACCAAGATTTC-3'; set 4 5'-CTAACCAACAATTGAGTTAATAGAC-3' and 5'-AGTAAACAAGGGTGTTTT-3'; for PB1: 5'-CAGCGGAAATGCTCGCAAAT-3' and 5'-TTGAGCTGTTGCTGGTCCAA-3') that amplify the region containing the PB1 M523I mutation and the complete HA gene segment. These fragments, approximately 500–600 nucleotides in length, were sequenced using the Roche 454 GS Junior sequencing platform. The fragment library was created for each sample according to the manufacturer's protocol without DNA fragmentation (GS FLX Titanium Rapid Library Preparation, Roche). The emulsion PCR (Amplification Method Lib-L) and GS Junior sequencing run were performed according to instructions of the manufacturer (Roche). Sequence reads from the GS Junior sequencing data were sorted by barcode and aligned to reference sequence A/Anhui/1/2013 using CLC Genomics software 6.0.2. Primers used to amplify the fragments were trimmed at the 3' and 5' ends of the sequence reads. For quality control, sequence reads were trimmed at the 3' end for Phred scores less than 30. The threshold for the detection of single nucleotide polymorphisms was manually set at 1% and 5%.

Model generation. A model of the structure of the HA of Anhui/1 was built by using MODELLER³² based upon the crystal structure of HA of H7N3 virus A/turkey/Italy/02 (Protein Data Bank code 1T18). The N123D and N149D mutations were introduced into the structure using the program Andante³³. Three-sugar glycans NeuAc α 2,6Gal β 1-4GlcNAc and NeuAc α 2,3Gal β 1-3GlcNAc were docked into the binding site of the Anhui/1 HA structure and that of the Anhui/1 (N123D/N149D) HA structure. Several strategies were then used to explore the docking of the glycan within the binding pocket. Alternative glycan conformations were produced by altering the phi angle of the glycosidic bond between the second and third sugars, and the exploration of alternative side-chain conformations of

amino acids within the binding pocket was performed by using a rotamer search³³. Lower energy structures were produced iteratively, until no further energy changes were seen. All simulations were performed using the University of Cambridge CAMGRID computing cluster³⁴.

Modified TRBC haemagglutination assay. Modified TRBC assays were performed as described previously³⁵. Briefly, all α 2,3-, α 2,6-, α 2,8- and α 2,9-linked sialic acids were removed from the surface of TRBCs by incubating 62.5 μ l of 1% TRBCs in PBS with 50 mU *Vibrio cholerae* NA (VCNA; Roche) in 8 mM calcium chloride at 37 °C for 1 h. Removal of sialic acids was confirmed by observation of complete loss of haemagglutination of the TRBCs by control influenza A viruses. Subsequently, resialylation was performed using 0.25 mU of α 2,3-(N)-sialyltransferase (COSMOBIO, bio-connect) or 12 mU of α 2,6-(N)-sialyltransferase (COSMOBIO, bio-connect) and 1.5 mM CMP-sialic acid (Sigma-Aldrich) at 37 °C in 75 μ l for 2 h to produce α 2,3-TRBC and α 2,6-TRBC, respectively. After a washing step, the TRBCs were re-suspended in PBS containing 1% bovine serum albumin to a final concentration of 0.5% TRBCs. Resialylation was confirmed by haemagglutination of viruses with known receptor specificity; recombinant viruses with six or seven gene segments of influenza virus A/Puerto Rico/8/1934 and the HA and NA of A/Vietnam/11/2004 H5N1 without the basic cleavage site or the HA of A/Netherlands/213/2003 H3N2. The receptor specificity of mutant viruses (recombinant viruses with seven gene segments of influenza virus A/Puerto Rico/8/1934 and the HA of Anhui/1 with or without the substitutions D123N and/or D149N) was tested by performing a standard haemagglutination assay with the modified TRBCs. In brief, serial twofold dilutions of virus in PBS were made in a 50 μ l volume; 50 μ l of 0.5% TRBCs were added, followed by incubation for 1 h at 4 °C before determining the haemagglutination titre.

Fusion assay. Influenza virus HA-induced cell fusion was tested in Vero-118 cells transfected with 5 μ g of pCAGGs-HA using Xtremegene transfection reagent (Roche). One day after transfection, cells were collected using trypsin-EDTA and plated in 6-well plates. The next morning, cells were washed and medium was replaced with IMDM medium containing 10 μ g ml⁻¹ of trypsin. After 1 h, cells were washed with PBS and exposed to PBS at pH 5.0, 5.2, 5.4, 5.6, 5.8, or 6.0 for 10 min at 37 °C. Subsequently, the PBS was replaced by IMDM supplemented with 10% FBS. Eighteen hours after the pH shock, cells were fixed using 80% ice-cold acetone, washed, and stained using a 20% Giemsa solution (Merck Millipore).

HA stability assay. The stability of HAs from the mutant viruses (recombinant viruses with seven gene segments of influenza virus A/Puerto Rico/8/1934 and the HA of Anhui/1 with or without the substitutions N123D and/or N149D and the HA of H5N1 A/Indonesia/5/2005 (INDO) with and without the substitution T318I) was evaluated by performing a thermostability assay. In short, viruses were diluted to 64 HA units per 25 μ l using PBS. The samples were incubated in a thermal cycler for 30 min at a temperature of 50 °C, 52 °C, 54 °C, 56 °C and 58 °C. Subsequently, the HA titre was determined by performing a haemagglutination assay using turkey erythrocytes.

Minigenome assay. A model viral RNA (vRNA), consisting of the firefly luciferase open reading frame flanked by the non-coding regions (NCRs) of segment 8 of influenza A virus, under the control of a T7 RNA polymerase promoter was used for minigenome assay³⁶. The reporter plasmid (0.5 μ g) was transfected into 293T cells in 6-well plates, along with 0.5 μ g of each of the pHW2000 plasmids encoding PB2, PB1, PA and NP, and 1 μ g of pAR3132 expressing T7 RNA polymerase and 0.02 μ g of the *Renilla* luciferase expression plasmid pRL (Promega) as an internal control. At 48 h after transfection, luminescence was measured using the Dual-Glo Luciferase Assay System (Promega) according to instructions of the manufacturer in a TECAN Infinite F200 machine (Tecan). Relative light units (RLU) were calculated as the ratio of Firefly and *Renilla* luciferase luminescences.

Pathology and immunohistochemistry. After fixation in 10% neutral-buffered formalin, tissues were embedded in paraffin, sectioned at 3 μ m, and stained with haematoxylin and eosin for the detection of histological lesions by light microscopy. For the detection of virus antigen by immunohistochemistry, tissues were stained with a monoclonal antibody against influenza A virus nucleoprotein as the primary antibody as described previously³⁷. After determining the cell types expressing viral antigen, the percentage of positively staining cells per tissue was estimated and ranked on an ordinal scale: 0, 0% of cells; 1, 1–25%; 2, 25–50%; 3, >50%.

28. Reed, L. J. & Munch, H. A simple method of estimating fifty percent endpoints. *Am. J. Hyg.* **27**, 493–497 (1938).
29. Nishiura, H., Yen, H. L. & Cowling, B. J. Sample size considerations for one-to-one animal transmission studies of the influenza A viruses. *PLoS ONE* **8**, e55358 (2013).
30. Hirst, G. K. Studies of antigenic differences among strains of influenza A by means of red cell agglutination. *J. Exp. Med.* **78**, 407–423 (1943).
31. Hoffmann, E., Stech, J., Guan, Y., Webster, R. G. & Perez, D. R. Universal primer set for the full-length amplification of all influenza A viruses. *Arch. Virol.* **146**, 2275–2289 (2001).
32. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).
33. Smith, R. E., Lovell, S. C., Burke, D. F., Montalvo, R. W. & Blundell, T. L. Andante: reducing side-chain rotamer search space during comparative modeling using environment-specific substitution probabilities. *Bioinformatics* **23**, 1099–1105 (2007).
34. Calleja, M. *et al.* Grid Tool Integration Within the eMinerals Project (Proc. UK e-Science All Hands Meeting, 2004).
35. Nobusawa, E., Ishihara, H., Morishita, T., Sato, K. & Nakajima, K. Change in receptor-binding specificity of recent human influenza A viruses (H3N2): a single amino acid change in hemagglutinin altered its recognition of sialyloligosaccharides. *Virology* **278**, 587–596 (2000).
36. de Wit, E. *et al.* Molecular determinants of adaptation of highly pathogenic avian influenza H7N7 viruses to efficient replication in the human host. *J. Virol.* **84**, 1597–1606 (2010).
37. van Riel, D., Rimmelzwaan, G. F., van Amerongen, G., Osterhaus, A. D. & Kuiken, T. Highly pathogenic avian influenza virus H7N7 isolated from a fatal human case causes respiratory disease in cats but does not spread systemically. *Am. J. Pathol.* **177**, 2185–2190 (2010).

Wapl is an essential regulator of chromatin structure and chromosome segregation

Antonio Tedeschi¹, Gordana Wutz¹, Sébastien Huet^{2,3,4}, Markus Jaritz¹, Annelie Wuensche², Erika Schirghuber^{1†}, Iain Finley Davidson¹, Wen Tang¹, David A. Cisneros^{1,5,6}, Venugopal Bhaskara^{1†}, Tomoko Nishiyama^{1†}, Alipasha Vaziri^{1,5,6}, Anton Wutz^{1†}, Jan Ellenberg² & Jan-Michael Peters¹

Mammalian genomes contain several billion base pairs of DNA that are packaged in chromatin fibres. At selected gene loci, cohesin complexes have been proposed to arrange these fibres into higher-order structures^{1–7}, but how important this function is for determining overall chromosome architecture and how the process is regulated are not well understood. Using conditional mutagenesis in the mouse, here we show that depletion of the cohesin-associated protein Wapl^{8,9} stably locks cohesin on DNA, leads to clustering of cohesin in axial structures, and causes chromatin condensation in interphase chromosomes. These findings reveal that the stability of cohesin–DNA interactions is an important determinant of chromatin structure, and indicate that cohesin has an architectural role in interphase chromosome territories. Furthermore, we show that regulation of cohesin–DNA interactions by Wapl is important for embryonic development, expression of genes such as *c-myc* (also known as *Myc*), and cell cycle progression. In mitosis, Wapl-mediated release of cohesin from DNA is essential for proper chromosome segregation and protects cohesin from cleavage by the protease separase, thus enabling mitotic exit in the presence of functional cohesin complexes.

Chromosome segregation depends on sister chromatid cohesion mediated by cohesin (reviewed in ref. 10), whose subunits Smc1 (also known as Smc1a), Smc3 and Scc1 (also known as Rad21 or Mcd1) embrace DNA strands as a ring¹¹. In mitotic prophase, cohesin is removed from chromosome arms by Wapl^{8,9}, which might open the Smc3–Scc1 interface^{10,12,13}, and in metaphase cohesin is removed from centromeres by separase, which opens the cohesin ring by cleaving Scc1 (refs 14, 15). Separase is essential for sister chromatid separation^{14,16–18}, but it is unclear whether Wapl is also required for this process, despite the fact that Wapl releases the bulk of cohesin from chromosomes.

Cohesin also associates with unreplicated DNA, presumably because cohesin has additional functions in gene regulation¹⁹. These have been proposed to depend on long-range chromatin interactions that are mediated by cohesin together with the transcription factors CCCTC-binding factor (CTCF) and mediator^{1–7}. How important cohesin-mediated chromatin interactions are for the structural and functional organization of mammalian genomes, and how these interactions are regulated to allow changes in chromatin structure and gene expression, are poorly understood. We therefore analysed whether Wapl is required for releasing cohesin from DNA not only in mitosis but also in interphase, and whether the regulation of cohesin–DNA interactions by Wapl is important for chromatin structure, gene regulation and chromosome segregation.

For this purpose, we generated mice in which the *Wapl* (also known as *Wapal*) alleles were replaced by ‘floxed’ alleles (*Wapl^{fl}*), which can be converted by Cre recombinase into *Wapl^Δ* ‘null’ alleles lacking exons 3 and 4 (called *Wapl^Δ* after germline transmission; Supplementary Fig. 1a, b and Methods). No *Wapl^{Δ/Δ}* mice were born (data not shown), indicating

that *Wapl* is essential for development. We therefore used mouse embryonic fibroblasts (MEFs), in which *Wapl^{fl}* is deleted after 4-hydroxytamoxifen (4-OHT) activation of the fusion protein between Cre and the oestrogen receptor (Cre-ER(T2)). When these cells, arrested in a quiescent state (G0) by serum starvation, were treated with 4-OHT, Wapl levels decreased over 7 days (Fig. 1a, b and Supplementary Fig. 1c, d). Fluorescence recovery after photobleaching (FRAP) analysis revealed that *Wapl* deletion increases the chromatin-residence time of cohesin by more than 20-fold: in quiescent *Wapl^{fl/fl}* MEFs, Scc1 tagged with green fluorescent protein (Scc1–GFP) bound to chromatin for 25 ± 9 min (mean ± s.d., *n* = 8), but in *Wapl^{Δ/Δ}* MEFs, 73 ± 9% of Scc1–GFP had a chromatin-residence time of 540 ± 240 min (*n* = 10; Fig. 1c and Supplementary Fig. 2). Consequently, most cohesin accumulated on chromatin (Fig. 1d). Wapl is therefore essential for the proper release of cohesin from chromatin.

Unexpectedly, immunofluorescence microscopy (IFM) experiments revealed that Wapl depletion changes cohesin localization: cohesin was located in most chromatin areas of quiescent *Wapl^{+/Δ}* MEFs, but in 80% of *Wapl^{Δ/Δ}* MEFs cohesin was enriched in elongated nuclear structures, which we call vermicelli (Italian for ‘little worms’; Fig. 1e and Supplementary Fig. 3). Vermicelli were also seen by live-cell imaging, ruling out fixation artefacts (Supplementary Fig. 4). CTCF was partially enriched in vermicelli, whereas histone H2B localization remained unchanged (Supplementary Fig. 5). Co-staining of chromosome 11 and cohesin by fluorescence *in situ* hybridization (FISH) and IFM, respectively, suggested that one vermicello is present per chromosome territory (*n* = 3) (Supplementary Fig. 6a, b). Labelling of individual chromosomes with Cy3–dUTP confirmed this (*n* = 31 chromosomes) (Fig. 1f, g and Supplementary Fig. 6c, d). These results indicate that in *Wapl*-deficient cells, cohesin is located in an axial chromosomal domain that might extend from one telomere to the other. Wapl depletion and stabilization of cohesin on DNA could either promote interactions between cohesin complexes, or enhance an effect that cohesin might have on chromatin structure, such as intra-chromatid loop formation, and lead to cohesin clustering indirectly (Fig. 1h, i).

Chromatin immunoprecipitation and Solexa DNA sequencing (ChIP-seq) showed that most cohesin and CTCF sites remained unchanged after Wapl depletion (Supplementary Fig. 7), suggesting that vermicelli are composed of cohesin sites that also exist in Wapl-proficient cells. We therefore analysed whether vermicelli are also present in wild-type cells. Although most cohesin had a fine-punctate distribution in chromatin, we observed short axial staining patterns in *Wapl^{+/+}* MEFs and HeLa cells. The axial cohesin staining overlapped only partially with H2B (data not shown), and became more prominent after *Wapl* RNA interference (RNAi; Supplementary Fig. 8), indicating that some cohesin is also located in axial domains in wild-type chromosomes.

¹Research Institute of Molecular Pathology (I.M.P.), Dr. Bohr-Gasse 7, 1030 Vienna, Austria. ²Cell Biology and Biophysics Unit, European Molecular Biology Laboratory (EMBL), Meyerhofstr. 1, Heidelberg 69117, Germany. ³CNRS, UMR 6290, Institut Génétique et Développement de Rennes, F-35043 Rennes, France. ⁴Université de Rennes 1, Université Européenne de Bretagne, Structure fédérative de recherche Biosit, Faculté de Médecine, F-35043 Rennes, France. ⁵Max F. Perutz Laboratories, University of Vienna, Austria, Dr. Bohr-Gasse 9, 1030 Vienna, Austria. ⁶Research Platform Quantum Phenomena & Nanoscale Biological Systems (QuNaBioS), University of Vienna, Dr. Bohr-Gasse 7-9, 1030 Vienna, Austria. [†]Present addresses: CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Lazarettgasse 14, AKH BT 25.3, 1090 Vienna, Austria (E.S.); Eucodis Bioscience GmbH, Campus Vienna Biocenter 2, 2. OG Viehmarktgassee 2A, 1030 Vienna, Austria (V.B.); Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan (T.N.); Institute of Molecular Health Sciences, ETH Zürich, Schafmattstrasse 22, 8093 Zürich, Switzerland (A.W.).

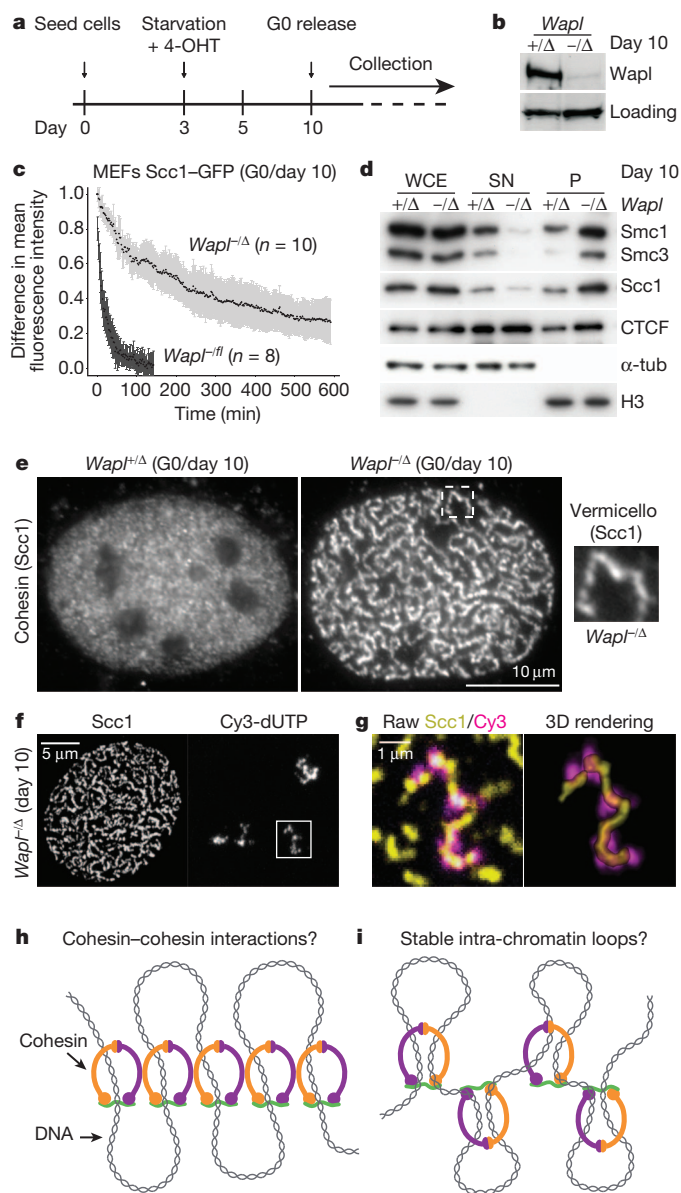


Figure 1 | Wapl depletion reveals arrangement of cohesin in axial chromosomal domains. **a**, Protocol for Wapl deletion by 4-OHT in G0-phase MEFs. **b**, Immunoblot of protein extracts from MEFs of indicated genotypes, treated as in **a**. Loading denotes unidentified protein cross-reacting with Wapl antibodies. **c**, FRAP analysis of MEFs expressing Scc1-GFP. The difference in fluorescence intensity between bleached and unbleached regions is plotted against time. Black dots denote mean values; error bars denote s.d. **d**, Immunoblot of whole-cell extracts (WCE), supernatant (SN) and pellet (P) fractions from MEFs. α -tub, α -tubulin. **e**, Microscopy images of non-preextracted MEFs stained for Scc1. Scale bar, 10 μ m. **f**, Maximum intensity projections of confocal z-stack images of MEFs in which chromosome territories were labelled with Cy3-dUTP, and stained for Scc1. Maximum projection was restricted to z-planes encompassing Cy3-dUTP signals. Scale bar, 5 μ m. **g**, Magnified images of regions boxed in **f**. Scale bar, 1 μ m. **h**, **i**, Alternative hypothetical models for how Wapl depletion might cause vermicello formation and chromatin compaction.

Wapl depletion also had major effects on chromatin structure: in FISH experiments, chromosome 11 occupied a significantly smaller median volume in Wapl^{-/-} MEFs (7×10^2 voxel) than in Wapl^{+/Δ} MEFs (8×10^2 voxel; Supplementary Fig. 6a, b). Giemsa staining showed that chromatin appeared more condensed in 78.5% of Wapl^{-/-} MEFs (Fig. 2a, b). A 'granularity index' analysis of 4',6-diamidino-2-phenylindole (DAPI)-stained cells (see Methods) confirmed this (Fig. 2c and Supplementary Fig. 6e). By contrast, no changes in the distribution

of heterochromatin protein 1a (HP1 α), histone H3 trimethylated on K9 (H3K9me3), the condensin subunit Smc2, and topoisomerase-II α and topoisomerase-II β could be detected in IFM and immunoblotting experiments (data not shown). Partial cohesin depletion by Scc1 RNAi reduced chromatin condensation in Wapl^{-/-} MEFs (Fig. 2). These observations indicate that Wapl deficiency causes chromatin compaction by stabilizing cohesin on DNA, and not indirectly. The compaction state of interphase chromatin, therefore, depends on the stability of cohesin–DNA interactions, which is controlled by Wapl.

When stimulated with serum, quiescent Wapl^{-/-} MEFs failed to proliferate (Fig. 3a). IFM experiments in which cells were released from quiescence in the continuous presence of bromodeoxyuridine (BrdU; Supplementary Fig. 9a, b) showed that fewer Wapl^{-/-} than Wapl^{+/Δ} MEFs incorporated BrdU. Wapl^{-/-} cells that had been able to enter S or G2 phase also contained residual Wapl (data not shown), indicating that cells fully depleted of Wapl do not replicate DNA. Degradation of cyclin-dependent kinase inhibitor p27^{Kip1} and cyclin A2 expression were reduced in Wapl^{-/-} MEFs (Supplementary Fig. 9c), indicating that Wapl is required for G1-phase progression. To test whether Wapl is also needed for DNA replication, we depleted Wapl from *Xenopus* egg extracts. This did not abrogate DNA replication (Supplementary Fig. 10), indicating that Wapl deficiency prevents S phase indirectly by interfering with exit from quiescence and/or G1 progression.

Because cell cycle progression depends on transcription, we tested whether Wapl is required for gene regulation. As serum stimulates proliferation in Wapl-proficient but not Wapl-deficient cells, we analysed quiescent cells to avoid indirect cell cycle effects (Supplementary Fig. 3). When we compared the levels of 21,169 RefSeq messenger RNAs between Wapl^{+/Δ} and Wapl^{-/-} MEFs using DNA microarrays, we found 1,152 differentially expressed genes (DEGs; 469 upregulated, 683 downregulated), including the cell cycle transcription factor *c-Myc*, the mRNA of which was reduced fourfold (Supplementary Fig. 11). Wapl^{-/-} MEFs in which *c-Myc* was expressed ectopically still did not proliferate, but entered the S and G2 phase more frequently (Supplementary Fig. 12), indicating that Wapl enables cell cycle progression in part by promoting *c-Myc* expression. Wapl depletion increased the mean Smc3-ChIP read density at the transcription start site (TSS) of many downregulated genes, including *c-myc*, and at non-TSS cohesin sites (Supplementary Fig. 13), indicating that Wapl depletion downregulates some genes by stabilizing cohesin on DNA. However, Wapl depletion decreased the mean Smc3-ChIP read density at the TSSs of most upregulated genes (Supplementary Fig. 14). How Wapl affects cohesin at these genes and increases their expression remains to be determined, but our results raise the possibility that changes in gene expression depend on the regulation of cohesin–DNA interactions by Wapl. Precedence for the regulatability of cohesin–DNA interactions comes from the observation that sister chromatid cohesion depends on inhibition of Wapl by sororin²⁰, which stabilizes cohesin on DNA²¹, and from the finding that Wapl-mediated cohesin release is activated in prophase⁹.

To investigate whether Wapl is required for mitosis, we analysed mitotic progression by IFM in Wapl^{-/-} MEFs containing sufficient Wapl to support one round of cell division (Fig. 3b and Supplementary Fig. 15a–d). In these cells, chromosome bridges occurred frequently in anaphase and telophase and many cells became bi-nucleated (Fig. 3c–e and Supplementary Fig. 15e–g). Although partial Scc1 depletion by RNAi reduced the chromosome bridge frequency (Fig. 3f), we were unable to detect Scc1 on these bridges (Fig. 3d and Supplementary Fig. 16a, c). The chromosome bridge phenotype therefore depends on cohesin, but may not be caused by an inability of Wapl-depleted cells to remove cohesin from chromosomes at anaphase onset. Instead, Wapl depletion might cause chromosome bridges by preventing the release of cohesin in early mitosis, which might be required for decatenation of sister chromatids^{22,23}. Consistently, at least one chromosome bridge in $54.5 \pm 14.9\%$ of Wapl-deficient MEFs contained the Bloom syndrome helicase (BLM) (Fig. 3g; triplicate experiments; n cells with bridges = 86), which has been implicated in de-catenation²⁴

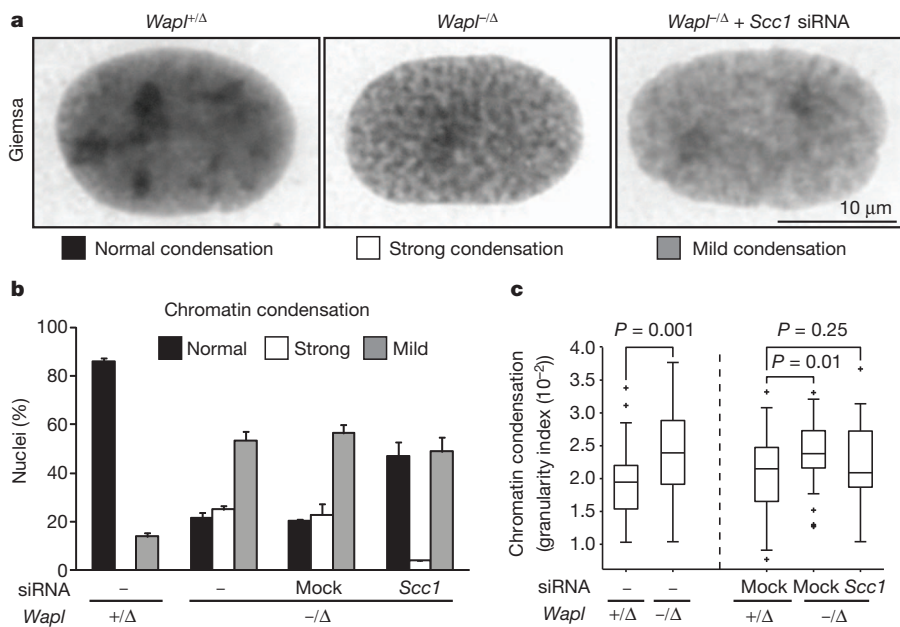


Figure 2 | Wapl controls chromatin structure by regulating cohesin–DNA interactions.

a, Microscopy images of Giemsa-stained interphase nuclei from MEFs of indicated genotypes. siRNA, short interfering RNA. Scale bar, 10 μm. **b**, Quantification of nuclear morphologies in cells from **a**, shown as mean and s.d. of two experiments ($n = 200$ per condition). **c**, Dense DAPI chromatin structures in MEFs were quantified by granularity index ($n > 40$; P values, Mann–Whitney U test). The boxes have lines at the lower quartile, median and upper quartile values.

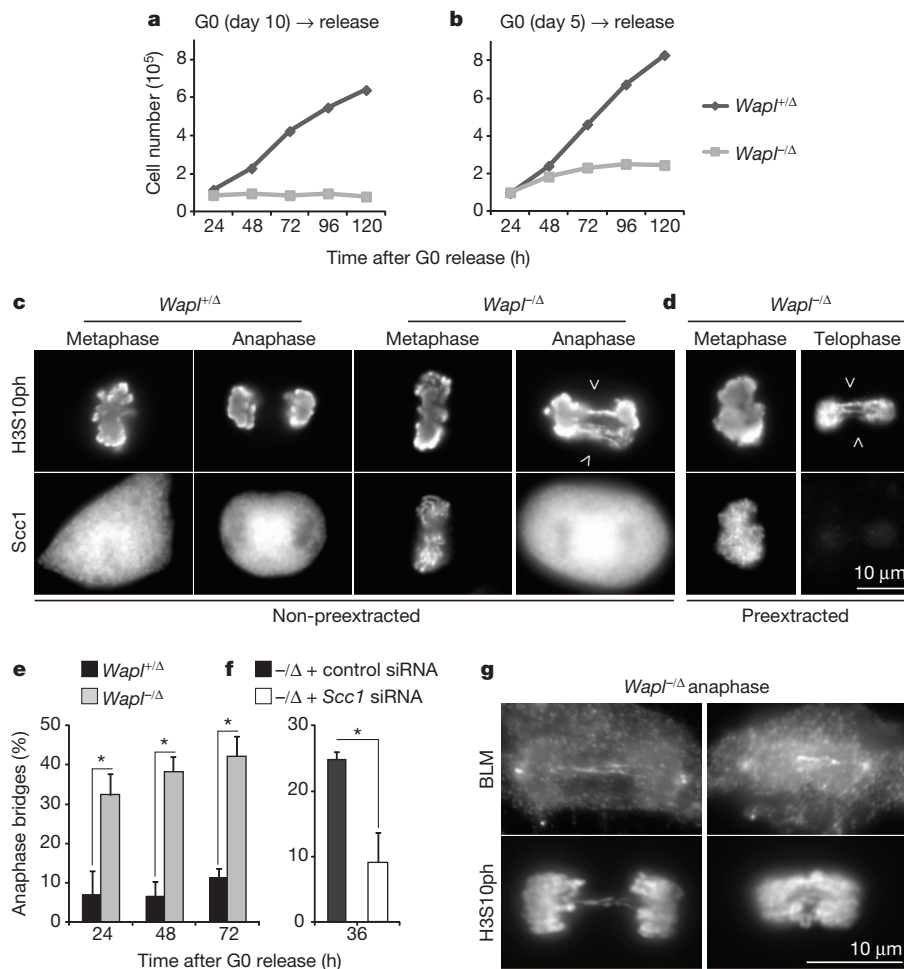


Figure 3 | Wapl is essential for cell cycle progression and proper chromosome segregation. **a**, **b**, Proliferation curves of MEFs obtained as in Fig. 1a. **c**, **d**, IFM images of metaphase, anaphase and telophase MEFs 72 h after G0 release as in **b**, either preextracted (**d**) or not (**c**), and co-stained for *Scc1* and histone H3 phosphorylated on Ser 10 (H3S10ph). Arrowheads indicate chromosome bridges. Scale bar, 10 μm. **e**, Quantification of chromosome

bridge frequency in MEFs from **c** and **d**, shown as the mean and s.d. of three experiments ($n \geq 100$ per condition; $*P < 0.05$, two-tailed Student's t -test). **f**, Quantification of chromosome bridge frequency in *Wapl*^{-/-} MEFs treated with *Scc1* or control siRNA, shown as the mean and s.d. of three experiments ($n \geq 156$ per condition; $P < 0.05$). **g**, IFM images of *Wapl*^{-/-} anaphase MEFs 48 h after G0 release (**b**), co-stained for BLM and H3S10ph. Scale bar, 10 μm.

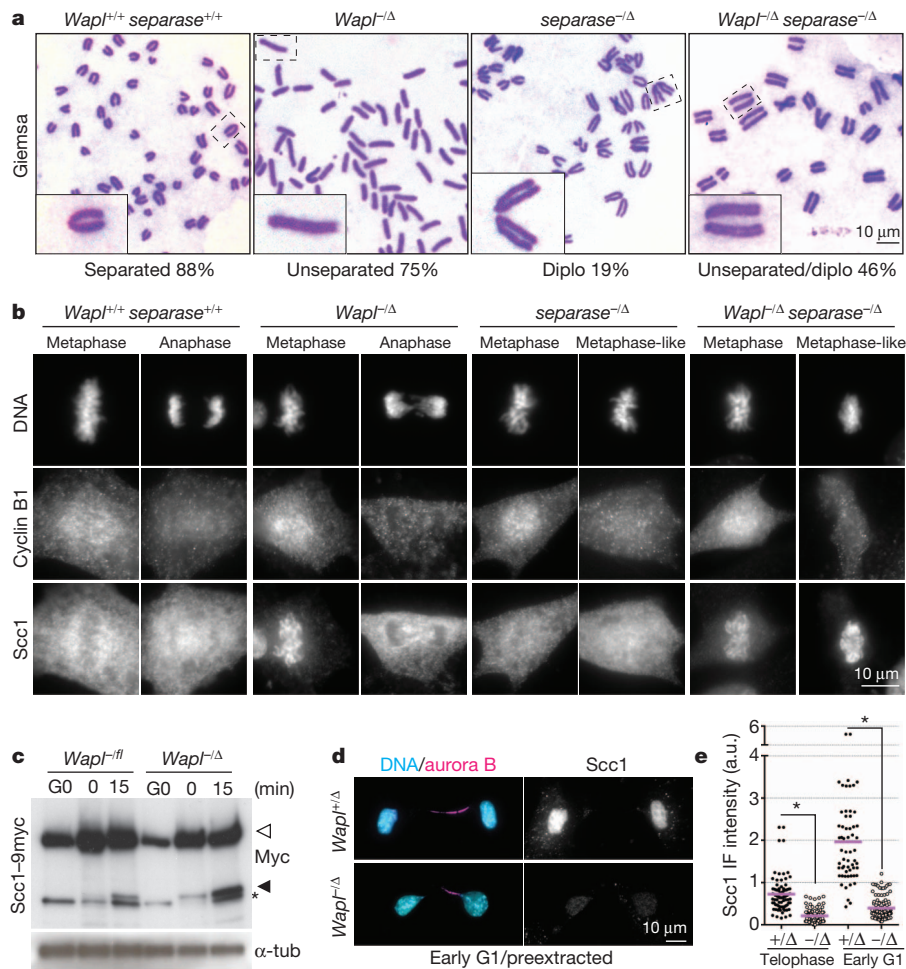


Figure 4 | The prophase pathway of cohesin dissociation protects cohesin from cleavage by separase. **a**, Microscopy images of Giemsa-stained prometaphase chromosomes from MEFs of indicated genotypes. Diplo, diplo-chromosomes ($n = 100$ prometaphase cells). Scale bar, $10\ \mu\text{m}$. **b**, IFM images of metaphase and anaphase MEFs co-stained for DAPI, Scc1 and cyclin B1. Scale bar, $10\ \mu\text{m}$. **c**, Immunoblot of extracts from MEFs expressing Scc1-9myc synchronized in G0, prometaphase (0 min) or anaphase (15 min). White and black

(for further characterization of these bridges, see Supplementary Fig. 17). Therefore, proper sister chromatid separation in anaphase depends on early mitotic release of cohesin from chromosomes by Wapl.

In $Wapl^{-/-}$ MEFs from which soluble proteins had not been pre-extracted, Scc1 was located on chromosomes in metaphase but cytoplasmically in anaphase ($n = 200$), indicating that Wapl deficiency delays cohesin release from prophase to anaphase onset (Fig. 3c, d). To test whether separase is required for releasing cohesin from chromosomes in Wapl-depleted cells, we analysed MEFs in which both the *Wapl* and *separase* (also known as *Espl1*) genes were deleted. Chromosome spread analysis confirmed that Wapl and separase deficiency prevent chromosome arm separation in prometaphase^{8,9} and centromere separation in anaphase, respectively, leading to diplo-chromosome formation owing to mitotic exit with unseparated sister chromatids in separase-depleted cells^{17,18}. Prometaphase $Wapl^{-/-}$ $separase^{-/-}$ cells showed a combination of both phenotypes, including diplo-chromosomes with unseparated arms (Fig. 4a and Supplementary Fig. 18a), indicative of defects in arm and centromere separation. IFM analysis revealed that in $Wapl^{-/-}$ $separase^{-/-}$ MEFs, sister chromatids remained connected and cohesin persisted on chromosomes even when cells exited mitosis, as measured by a decrease in cyclin B1 levels and re-localization of aurora B from centromeres to mid-spindles (Fig. 4b and Supplementary Fig. 18b–d). Separase is therefore required for releasing cohesin from both centromeres and chromosome arms in Wapl-depleted cells.

arrowheads and asterisk indicate full-length Scc1-9myc, an anaphase-specific cleavage product of Scc1-9myc, and a cross-reacting band, respectively. **d**, IFM images of preextracted early G1-phase MEFs co-stained for DAPI, aurora B and Scc1. Scale bar, $10\ \mu\text{m}$. **e**, Quantification of Scc1 immunofluorescence (IF) signals from MEFs analysed as in **d**. Magenta bars denote means of three experiments. Dots denote single data points ($n > 37$ per condition; $*P < 0.0001$, two-tailed Student's *t*-test). $+/\Delta$ and $-/\Delta$ indicate $Wapl^{+/+}$ and $Wapl^{-/-}$, respectively.

We noticed that separase cleaved more Scc1 in Wapl-depleted cells than in non-depleted cells, presumably because separase cleaves chromosome-bound but not soluble cohesin^{15,25} (Fig. 4c; note that these cells also contained uncleaved Scc1, possibly because cells had been treated only for two days with 4-OHT and thus contained residual Wapl, and/or because in these experiments only 35% of MEFs entered anaphase after release from the prometaphase arrest; Supplementary Fig. 16a–d). This suggested that Wapl-deficient cells might exit mitosis with less intact cohesin and can, therefore, not load as much cohesin onto chromatin as normally in the next cell cycle. Indeed, IFM analysis showed that significantly less chromatin-bound Scc1 was detectable in $Wapl^{-/-}$ than in $Wapl^{+/+}$ MEFs in telophase and G1 phase (Fig. 4d, e and Supplementary Fig. 16e; note that in this experiment Wapl was depleted more efficiently than in Fig. 4c by treating cells for 7 days with 4-OHT, presumably causing more Scc1 cleavage than in Fig. 4c). These results show that the prophase pathway of cohesin dissociation allows mitotic exit in the presence of functional cohesin, enabling it to function early in the next cell cycle.

Our findings that Wapl depletion ‘locks’ cohesin on chromatin in a stably bound state and alters chromatin compaction to a degree that can be seen by light microscopy, indicate that the stability of cohesin–DNA interactions is an important determinant of chromatin structure, perhaps because the residence time of cohesin on DNA determines whether and for how long intra-chromatid loops can be formed (Fig. 1h, i). The

role of Wapl in chromatin structure may be evolutionarily conserved, as alterations in chromosome morphology have been observed in fruit-fly and yeast *Wapl* (*wapl* and *RAD61*, respectively) mutants^{26–28}. Our observation that cohesin complexes form axial structures (vermicelli) in interphase chromosomes of Wapl-depleted cells, and to a lesser degree in wild-type cells, indicates that cohesin has an architectural role in the organization of interphase chromosome territories. Similar roles have been proposed for cohesin and the related condensin complexes in meiotic and mitotic chromosomes, respectively^{29,30}. Because cohesin-related complexes exist in all kingdoms of life, the architectural role of cohesin that we propose here for interphase chromosomes may represent an ancient function that may have helped to organize DNA in chromosomes already before nucleosomes existed.

METHODS SUMMARY

Generation of conditional *Wapl* allele. The conditional *Wapl* allele (*Wapl^{fl}*) was generated by inserting *loxP* sites into introns 2 and 4 to allow Cre-mediated excision of exons 3 and 4 (Supplementary Fig. 1). Correctly targeted embryonic stem cell clones were identified by Southern blotting of EcoRI- and SacI-digested genomic DNA to confirm the presence of 5' and 3' *loxP* sites, respectively (data not shown). **Cell culture.** Immortalized MEFs (iMEFs) were used for the fractionation experiments in Fig. 1d, the nucleotide labelling experiment in Fig. 1f, and for experiments in which tagged versions of Scc1 were stably expressed. All the other experiments were performed in primary MEFs from embryonic day (E)13.5 embryos. MEFs and iMEFs were cultured as described¹⁸.

To synchronize MEFs or iMEFs in quiescence and to activate CreER(T2), cells were first cultured to confluence and then serum starved for 7 or 2 days (corresponding to day 10 and 5 in Fig. 1a, respectively) with optiMEM media (Invitrogen) supplemented with 2% charcoal/dextran-treated serum (Hyclone), 100 U ml⁻¹ penicillin (Sigma), 100 µg ml⁻¹ streptomycin (Sigma) and 0.5 µM 4-OHT (Sigma, 10 mM stock in ethanol). Cells were split and released in fresh medium (10% serum) for cell cycle analysis.

Cells were enriched in prometaphase of mitosis by nocodazole (Sigma) treatment for 5 h at 300 ng ml⁻¹ (Fig. 4a) or 200 ng ml⁻¹ (Fig. 4c). Cells were either continuously labelled (Supplementary Fig. 9a) or pulsed (2 h) (Supplementary Fig. 12f) with 40 µM BrdU (Sigma).

Full Methods and any associated references are available in the online version of the paper.

Received 23 November 2012; accepted 16 July 2013.

Published online 25 August 2013.

- Hadjir, S. *et al.* Cohesins form chromosomal cis-interactions at the developmentally regulated *IFNG* locus. *Nature* **460**, 410–413 (2009).
- Mishiro, T. *et al.* Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J.* **28**, 1234–1245 (2009).
- Nativio, R. *et al.* Cohesin is required for higher-order chromatin conformation at the imprinted *IGF2-H19* locus. *PLoS Genet.* **5**, e1000739 (2009).
- Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
- Seitan, V. C. *et al.* A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature* **476**, 467–471 (2011).
- DeMare, L. E. *et al.* The genomic landscape of cohesin-associated chromatin interactions. *Genome Res.* <http://dx.doi.org/10.1101/gr.156570.113> (23 May 2013).
- Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
- Gandhi, R., Gillespie, P. J. & Hirano, T. Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. *Curr. Biol.* **16**, 2406–2417 (2006).
- Kueng, S. *et al.* Wapl controls the dynamic association of cohesin with chromatin. *Cell* **127**, 955–967 (2006).
- Peters, J. M., Tedeschi, A. & Schmitz, J. The cohesin complex and its roles in chromosome biology. *Genes Dev.* **22**, 3089–3114 (2008).
- Haering, C. H., Farcas, A. M., Arumugam, P., Metson, J. & Nasmyth, K. The cohesin ring concatenates sister DNA molecules. *Nature* **454**, 297–301 (2008).
- Chan, K. L. *et al.* Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. *Cell* **150**, 961–974 (2012).
- Buheitel, J. & Stemmann, O. Prophase pathway-dependent removal of cohesin from human chromosomes requires opening of the Smc3-Scc1 gate. *EMBO J.* **32**, 666–676 (2013).
- Uhlmann, F., Lottspeich, F. & Nasmyth, K. Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. *Nature* **400**, 37–42 (1999).

- Waizenegger, I. C., Hauf, S., Meinke, A. & Peters, J. M. Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. *Cell* **103**, 399–410 (2000).
- Hauf, S., Waizenegger, I. C. & Peters, J. M. Cohesin cleavage by separase required for anaphase and cytokinesis in human cells. *Science* **293**, 1320–1323 (2001).
- Kumada, K. *et al.* The selective continued linkage of centromeres from mitosis to interphase in the absence of mammalian separase. *J. Cell Biol.* **172**, 835–846 (2006).
- Wirth, K. G. *et al.* Separase: a universal trigger for sister chromatid disjunction but not chromosome cycle progression. *J. Cell Biol.* **172**, 847–860 (2006).
- Seitan, V. C. & Merckenschlager, M. Cohesin and chromatin organisation. *Curr. Opin. Genet. Dev.* **22**, 93–100 (2012).
- Nishiyama, T. *et al.* Sororin mediates sister chromatid cohesion by antagonizing Wapl. *Cell* **143**, 737–749 (2010).
- Schmitz, J., Watrin, E., Lenart, P., Mechtler, K. & Peters, J. M. Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. *Curr. Biol.* **17**, 630–636 (2007).
- Wang, L. H., Mayer, B., Stemmann, O. & Nigg, E. A. Centromere DNA decatenation depends on cohesin removal and is required for mammalian cell division. *J. Cell Sci.* **123**, 806–813 (2010).
- Farcas, A. M., Ulucak, P., Helmhart, W. & Nasmyth, K. Cohesin's concatenation of sister DNAs maintains their intertwining. *Mol. Cell* **44**, 97–107 (2011).
- Chan, K. L. & Hickson, I. D. New insights into the formation and resolution of ultra-fine anaphase bridges. *Semin. Cell Dev. Biol.* **22**, 906–912 (2011).
- Sun, Y. *et al.* Separase is recruited to mitotic chromosomes to dissolve sister chromatid cohesion in a DNA-dependent manner. *Cell* **137**, 123–132 (2009).
- Verni, F., Gandhi, R., Goldberg, M. L. & Gatti, M. Genetic and molecular analysis of wings apart-like (*wapl*), a gene controlling heterochromatin organization in *Drosophila melanogaster*. *Genetics* **154**, 1693–1710 (2000).
- Guacci, V. & Koshland, D. Cohesin-independent segregation of sister chromatids in budding yeast. *Mol. Biol. Cell* **23**, 729–739 (2012).
- Lopez-Serra, L., Lengronne, A., Borges, V., Kelly, G. & Uhlmann, F. Budding yeast Wapl controls sister chromatid cohesion maintenance and chromosome condensation. *Curr. Biol.* **23**, 64–69 (2013).
- Pezzi, N. *et al.* STAG3, a novel gene encoding a protein involved in meiotic chromosome pairing and location of STAG3-related genes flanking the Williams–Beuren syndrome deletion. *FASEB J.* **14**, 581–592 (2000).
- Ono, T. *et al.* Differential contributions of condensin I and condensin II to mitotic chromosome architecture in vertebrate cells. *Cell* **115**, 109–121 (2003).

Supplementary Information is available in the online version of the paper.

Acknowledgements We dedicate this paper to the memory of B. Peters, who performed the first experiments on Wapl in our laboratory. We are grateful to K. Aumayr, O. F.-Capetillo, T. Hoffmann, M. E. Idarraga-Amado, S. Kueng, T. Kulcsar, M. Leeb, P. Pasierbek, D. Santamaria, G. Schmauss, A. Souabni, H. Tkadletz, K. Wendt and members of the Peters laboratory for discussions and assistance, J. Hutchins for suggesting the term vermicelli, K. Nasmyth for the *separase* mouse model, and M. Barbacid, T. Cremer, T. Hirano, T. Jenuwein, M. Malumbres and J. Zuber for reagents. T.N. was supported by the European Molecular Biology Organization (EMBO) and the Japanese Society for the Promotion of Science (JSPS). S.H. was supported by funds from the Agence National de la Recherche (JCJC-SVSE2-2011, ChromaTranscript project) and the European Union (FP7-PEOPLE-2011-CIG, ChromaTranscript project). D.A.C. was supported by MFPL VIPS Program (BMWF and City of Vienna). A.V. acknowledges financial support by the Vienna Science and Technology Fund (WTF) project VRG10-11, the Research Platform Quantum Phenomena and Nanoscale Biological Systems (QuNaBioS) and by Boehringer Ingelheim. Research in the laboratory of J.-M.P. is supported by Boehringer Ingelheim, the Austrian Science Fund (FWF special research program SFB F34 'Chromosome Dynamics', and Wittgenstein award Z196-B20), the Austrian Research Promotion Agency (FFG, Laura Bassi Center for Optimized Structural Studies), the Vienna Science and Technology Fund (WTF LS09-13), and the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 241548 (MitoSys).

Author Contributions Experiments were designed and data interpreted by A.T., G.W., S.H., M.J., J.E. and J.-M.P. A.T. generated *Wapl^{fl}* embryonic stem cells, established and maintained the mouse colony required for this study, and carried out chromosome segregation, transcriptome and ChIP-seq studies. A.T. and G.W. analysed cohesin localization by IFM, and cleavage of Scc1 in anaphase MEFs. G.W. generated MEFs expressing Scc1-LAP and Scc1-9myc, and carried out cohesin IFM experiments in telophase and G1-phase MEFs. G.W. and V.B. performed chromatin fractionation experiments. G.W. and T.N. performed cleavage assays of Scc1 with *Xenopus* egg extracts. A.T., G.W. and W.T. analysed cell cycle progression of MEFs. S.H. performed DNA granularity analysis and Cy3-dUTP labelling of DNA. M.J. carried out bioinformatic analyses. A.Wue. performed cohesin FRAP experiments. E.S. performed FISH experiments. A.Wut. supervised the work of A.T. with embryonic stem cells, and participated in designing the *Wapl* targeting strategy. I.F.D. performed replication assays with *Xenopus* egg extracts. D.A.C. performed live imaging of cohesin vermicelli under the guidance of A.V. A.T. and J.-M.P. wrote the manuscript.

Author Information ChIP-seq and microarray data have been deposited in the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>) under accession code GSE41603. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.-M.P. (jan-michael.peters@imp.ac.at).

METHODS

Generation of conditional *Wapl* allele. The conditional *Wapl* allele (*Wapl*^{f/f}) was generated by inserting *loxP* sites into introns 2 and 4 to allow Cre-mediated excision of exons 3 and 4 (Supplementary Fig. 1). Correctly targeted embryonic stem-cell clones were identified by Southern blotting of EcoRI- and SacI-digested genomic DNA to confirm the presence of 5' and 3' *loxP* sites, respectively (data not shown).

Mice. The conditional and null *Wapl* mice were generated by crossing with FLPe mice³¹ and MORE mice³², respectively. Separase mice¹⁸, Rosa26CreER(T2) mice³³, and H2B-GFP transgenic mice³⁴ have been described. Mice were hybrids between C57BL/6 and 129 strains.

Cell culture. Immortalized MEFs (iMEFs) were used for the fractionation experiments in Fig. 1d, the nucleotide labelling experiment in Fig. 1f, and for experiments in which tagged versions of Scc1 were stably expressed. All other experiments were performed in primary MEFs from embryonic day (E)13.5 embryos. MEFs and iMEFs were cultured as described¹⁸. *Wapl*^{f/f} Rosa26^{ERCre (T2)/+} iMEFs stably expressing mouse *Scc1* (cDNA) tagged with nine Myc epitopes (Scc1-9myc), eGFP (Scc1-eGFP) or Dronpa (Scc1-Dronpa)¹⁸, or mouse Scc1 (BAC RP23-375K15) fused to the LAP tag (Scc1-LAP)³⁵ were generated as described.

To synchronize MEFs or iMEFs in quiescence and to activate CreER(T2), cells were first cultured to confluence and then serum starved for 7 or 2 days (corresponding to day 10 and 5 in Fig. 1a, respectively) with optiMEM media (Invitrogen) supplemented with 2% charcoal/dextran-treated serum (Hyclone), 100 U ml⁻¹ penicillin (Sigma), 100 µg ml⁻¹ streptomycin (Sigma) and 0.5 µM 4-OHT (Sigma, 10 mM stock in ethanol). Cells were split and released in fresh medium (10% serum) for cell cycle analysis.

Cells were enriched in prometaphase of mitosis by nocodazole (Sigma) treatment for 5 h at 300 ng ml⁻¹ (Fig. 4a) or 200 ng ml⁻¹ (Fig. 4c). Cells were either continuously labelled (Supplementary Fig. 9a) or pulsed (2 h) (Supplementary Fig. 12f) with 40 µM BrdU (Sigma).

MEFs used in Fig. 4a were infected with AdCre-expressing virus¹⁸ and collected 48 h after infection for chromosome spread and Giemsa staining³⁶, as described. Plasmids for c-Myc rescue experiments (Supplementary Fig. 12) have been described previously³⁷.

Nucleotide labelling of interphase chromosome territories. To incorporate Cy3-dUTP in iMEFs, cells were grown on Lab-Tek chambered coverglass and synchronized at the G1/S transition by 2 days serum starvation followed by 42 h incubation in growing medium containing 1 µg ml⁻¹ aphidicolin (Sigma-Aldrich). Immediately after aphidicolin release, the cell layer, bathed with growing medium containing 10 µM Cy3-dUTP (GE Healthcare), was scratched using the tip of a syringe needle to allow nucleotide loading³⁸ and integration to the DNA during replication. After 4 days in growing medium and owing to the successive cell divisions, each nucleus contained only a few labelled chromosomes, each of them appearing as a dense group of fluorescent spots by light microscopy as described³⁹. Cells were then serum starved in presence of 4-OHT for 7 days and imaged.

Antibodies. Most of the antibodies have been used at 1 µg ml⁻¹, except where otherwise indicated.

Immunofluorescence. Rabbit anti-myc (myc (527), Peters laboratory ID A957)¹⁵, mouse anti-myc (myc (9E10), Peters laboratory ID A668)¹⁵, mouse anti-BrdU (BD 347580), mouse anti-BrdU (Caltag MD5110, 1:500), rabbit anti-Wapl (Afr266-3L, a gift from T. Hirano, 1:250), chicken anti-GFP (Abcam ab13970), rabbit anti-CTCF (CTCF-35, Peters laboratory ID A992), rabbit anti-cyclin B1 (Cell Signaling 4138S), mouse anti-aurora B (BD 611083), mouse anti-Scc1 (UPSTATE 05-908), rabbit anti-Scc1 (Scc1 (623) Peters laboratory ID A900)¹⁵, rabbit anti-H3S10ph (Millipore 06570, 1:500), mouse anti-H3S10ph (Cell Signaling 9706L, 1:100), rabbit anti-BLM (Abcam ab476, 1:200), rat anti-RPA32 (Cell Signaling 4EA, 1:200) and Alexa fluorophore (488, 568, 633)-conjugated antibodies were used (Molecular Probes, 1:1,000).

Western blotting. Rabbit anti-Wapl (hWapl (4748), Peters laboratory ID A960; hWapl (4749), Peters laboratory ID A961)⁹, goat anti-H3 (Santa Cruz sc-8654), rabbit anti-cyclin A (Santa Cruz sc-596), mouse anti-Scc1 (UPSTATE 05-908), rabbit anti-Smc3 (Bethyl A300-060A), rabbit anti-Smc1 (Bethyl A300-055A), mouse anti-tubulin (Sigma B512), rabbit anti-p27 (a gift from M. Barbacid, 1:250) and rabbit anti-CTCF (CTCF-35, Peters laboratory ID A992) were used.

Chromatin immunoprecipitation. Rabbit anti-Smc3 (Bethyl A300-060A)⁴⁰ and rabbit anti-CTCF (UPSTATE 07-729)⁴⁴ have been described. Methods for sub-cellular fractionation and western blotting⁹ have been described.

Immunofluorescence and immuno-FISH. Cells were grown on 22-mm glass coverslips (Menzel), fibronectin-coated slides (BD Biosciences), Lab-Tek chambered coverglass (Nunc), or spun onto glass slides (Menzel) using a Cytospin centrifuge (Shandon), fixed with 4% paraformaldehyde (PFA) and stained as described³⁶. Where indicated, cells were extracted using 0.1% Triton X-100 for 2 min before fixation. BrdU staining was performed as described⁴² except that cells were fixed

with 4% PFA instead of 70% ethanol. RPA32 and BLM staining was performed as described⁴³.

For immuno-FISH experiments, MEFs were fixed with 2% PFA and stained with a mouse anti-Scc1 antibody (UPSTATE 05-908) and Alexa 488 secondary antibody as described³⁶. MEFs were then fixed with 1% PFA and hybridized with mouse paint specific for mouse chromosome 11 labelled with a red emitting fluorochrome (Metasystems) according to the manufacturer's instruction. DNA was counterstained with 2 µg ml⁻¹ DAPI (Roche Diagnostics). Slides were mounted using Vectashield Mounting medium (Vector Laboratories).

Microscopy and image analysis. Epifluorescence microscopy was performed with an Axioplan 2 microscope (Zeiss) using a CoolSnap HQ camera (Photometrics) and images were processed with MetaMorph (Universal Imaging).

FRAP imaging and analysis was carried out as described⁴⁴. Cells were imaged with a Zeiss 780 confocal microscope using a Zeiss Fluor 40×/1.3 oil objective. After capture of a pre-bleach image, half of the nucleus was bleached followed by time-lapse imaging with 2-min time resolution over 2 h (control cells) and 10 h (Wapl-depleted cells). For each time point, measurements were taken in user-defined regions. iMEFs expressing Scc1-LAP were used in these experiments.

Single-point confocal laser scanning microscopy was performed with Zeiss 510 (Supplementary Fig. 5) and Zeiss LSM780 (Fig. 1f, g) with a plan-apochromat 63×/1.4 oil objective. Different fluorescence signals were collected in multitrack mode set up for three channels to fit with the fluorescence spectrum of the respective dyes. Colocalization of Scc1 and H2B or CTCF was performed with ImageJ⁴⁵. The association of cohesin vermicelli with dUTP-Cy3-labelled DNA is described in Supplementary Information. Confocal spinning disk microscopy (Supplementary Fig. 6a) was performed using Ultraview run under Velocity software with a plan-apochromat 63×/1.4 oil objective (Perkin Elmer). To avoid cross-talk of fluorophores, channels were acquired sequentially using bandpass filters covering respective emission spectra of used fluorophores. Deconvolution was performed using Huygens Professional (SVI) software using classic maximum likelihood estimation (CMLE) algorithm.

Live-cell imaging was performed in a Zeiss TIRF 3 microscope used in widefield mode at 37 °C using CO₂-independent medium (Invitrogen). To image Scc1-Dronpa and H2B-PAmCherry, cells were illuminated with a 405-nm laser for 5–10 s to activate its fluorescent state. Afterwards, cells could be imaged using a 488-nm and 561-nm laser for Scc1-Dronpa and H2B-PAmCherry, respectively. Hoechst and Scc1-eGFP were imaged using a 405-nm and 488-nm laser, respectively.

RNA interference. Transfection of control GL2 siRNA (Ambion), human WAPL siRNA (Ambion)⁹ and mouse Scc1 siRNA (Ambion, 72658) was performed with Lipofectamine RNAi MAX (Invitrogen) according to manufacturer's instruction. All siRNAs were used at a final concentration of 100 nM. MEFs in Fig. 2 were released from quiescent (day 10 in Fig. 1a) into medium containing the indicated siRNA. Cells were then analysed 48 h after transfection (Scc1 levels measured by quantitative IFM were reduced to 57 ± 13.4% (mean ± s.d. of two experiments)). **ChIP-seq data analysis.** Cohesin and CTCF ChIP was performed as described⁴¹, except that samples were sonicated for 12 × 60-s pulses with a Covaris sonicator, and the DNA was purified with Phase lock gel tubes (5PRIME). ChIP-seq sample preparation, sequencing and data analysis are described in Supplementary Information.

Transcriptome microarray analysis. RNAs were extracted from MEFs with the RNeasy Mini Kit (QIAGEN). Microarray hybridization and data analysis are described in Supplementary Information. Methods for quantitative reverse transcriptase PCR (qRT-qPCR)⁴¹ have been described.

Statistics. Statistical analyses were carried out using the GraphPad Prism software package (GraphPad Software). Statistical analyses for transcriptome and ChIP-seq experiments are described in Supplementary Information.

- Rodríguez, C. I. *et al.* High-efficiency deleter mice show that FLPe is an alternative to Cre-*loxP*. *Nature Genet.* **25**, 139–140 (2000).
- Tallquist, M. D. & Soriano, P. Epiblast-restricted Cre expression in MORE mice: a tool to distinguish embryonic vs. extra-embryonic gene function. *Genesis* **26**, 113–115 (2000).
- Seibler, J. *et al.* Rapid generation of inducible mouse mutants. *Nucleic Acids Res.* **31**, e12 (2003).
- Hadjantonakis, A. K. & Papaioannou, V. E. Dynamic *in vivo* imaging and cell tracking using a histone fluorescent protein fusion in mice. *BMC Biotechnol.* **4**, 33 (2004).
- Poser, I. *et al.* BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nature Methods* **5**, 409–415 (2008).
- Hauf, S. *et al.* The small molecule Hesperadin reveals a role for Aurora B in correcting kinetochore-microtubule attachment and in maintaining the spindle assembly checkpoint. *J. Cell Biol.* **161**, 281–294 (2003).
- Zuber, J. *et al.* RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478**, 524–528 (2011).

38. Schermelleh, L., Solovej, I., Zink, D. & Cremer, T. Two-color fluorescence labeling of early and mid-to-late replicating chromatin in living cells. *Chromosome Res.* **9**, 77–80 (2001).
39. Zink, D. *et al.* Structure and dynamics of human interphase chromosome territories in vivo. *Hum. Genet.* **102**, 241–251 (1998).
40. Stedman, W. *et al.* Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J.* **27**, 654–666 (2008).
41. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
42. Bugler, B., Schmitt, E., Aressy, B. & Ducommun, B. Unscheduled expression of CDC25B in S-phase leads to replicative stress and DNA damage. *Mol. Cancer* **9**, 29 (2010).
43. Cerqueira, A. *et al.* Overall Cdk activity modulates the DNA damage response in mammalian cells. *J. Cell Biol.* **187**, 773–780 (2009).
44. Gerlich, D., Koch, B., Dupeux, F., Peters, J. M. & Ellenberg, J. Live-cell imaging reveals a stable cohesin–chromatin interaction after but not before DNA replication. *Curr. Biol.* **16**, 1571–1578 (2006).
45. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).

Two replication fork maintenance pathways fuse inverted repeats to rearrange chromosomes

Lingchuan Hu^{1*}, Tae Moon Kim^{1*}, Mi Young Son¹, Sung-A Kim¹, Cory L. Holland¹, Satoshi Tateishi², Dong Hyun Kim¹, P. Renee Yew¹, Cristina Montagna³, Lavinia C. Dumitrache^{1†} & Paul Hasty¹

Replication fork maintenance pathways preserve chromosomes, but their faulty application at nonallelic repeats could generate rearrangements causing cancer, genomic disorders and speciation^{1–3}. Potential causal mechanisms are homologous recombination and error-free postreplication repair (EF-PRR). Homologous recombination repairs damage-induced DNA double-strand breaks (DSBs) and single-ended DSBs within replication. To facilitate homologous recombination, the recombinase RAD51 and mediator BRCA2 form a filament on the 3' DNA strand at a break to enable annealing to the complementary sister chromatid⁴ while the RecQ helicase, BLM (Bloom syndrome mutated) suppresses crossing over to prevent recombination⁵. Homologous recombination also stabilizes^{6,7} and restarts^{8,9} replication forks without a DSB^{10,11}. EF-PRR bypasses DNA incongruities that impede replication by ubiquitinating PCNA (proliferating cell nuclear antigen) using the RAD6–RAD18 and UBC13–MMS2–RAD5 ubiquitin ligase complexes¹². Some components are common to both homologous recombination and EF-PRR such as RAD51 and RAD18^{13,14}. Here we delineate two pathways that spontaneously fuse inverted repeats to generate unstable chromosomal rearrangements in wild-type mouse embryonic stem (ES) cells. Gamma-radiation induced a BLM-regulated pathway that selectively fused identical, but not mismatched, repeats. By contrast, ultraviolet light induced a RAD18-dependent pathway that efficiently fused mismatched repeats. Furthermore, TREX2 (a 3'→5' exonuclease) suppressed identical repeat fusion but enhanced mismatched repeat fusion, clearly separating these pathways. TREX2 associated with UBC13 and enhanced PCNA ubiquitination in response to ultraviolet light, consistent with it being a novel member of EF-PRR. RAD18 and TREX2 also suppressed replication fork stalling in response to nucleotide depletion. Interestingly, replication fork stalling induced fusion for identical and mismatched repeats, implicating faulty replication as a causal mechanism for both pathways.

The identical and mismatched repeat reporters (IRR and MRR, Fig. 1a, b) were designed to investigate pathways that rearrange chromosomes through repeat fusion. Both reporters contain a 313-base-pair major satellite repeat (MSR) at each junction of an inversion in *miniHPRT*. These repeats are indirect so repeat fusion restores *miniHPRT* to enable survival in hypoxanthine, aminopterin, thymidine (HAT)-selection media by a potential mechanism shown in Fig. 1c. The only difference between these reporters is that the MRR 3' repeat contains seven mismatches with the longest contiguous homology being 67 bases. The IRR and MRR were stably transfected into wild-type AB2.2 and IB10 ES cells. About the same number of HAT-resistant colonies spontaneously grew for both reporters (Fig. 1d, $P > 0.85$, Student's *t*-test), indicating that spontaneous repeat fusion occurred in wild-type cells.

The fused 5' repeat for the MRR was sequenced to determine the switch location (Fig. 1e, Extended Data Fig. 1). Strand exchange in fission yeast predominantly occurred at the palindrome centre after replication forks were induced to stall, an event called a U turn⁹. We

found 6 of 14 switches had this U-turn at the base of a putative hairpin (all green), whereas two occurred at the apex (all orange) and six occurred in the stem (green-orange). Thus, strand exchange occurred at multiple locations.

It is possible that the switched strand replicated to the telomere, forming a dipericentric (Fig. 1c). Two-colour fluorescence *in situ* hybridization (FISH) was performed on clones with the IRR and MRR using a pericentromeric and telomeric probe. Dipericentrics and chromosomes with extra pericentromeres and telomeres (EPTs)¹⁵ were observed for cells with both reporters (Extended Data Fig. 2a and Extended Data Tables 1 and 2). EPTs seemed unstable because the pericentromere number and location varied between metaphase spreads from the same clone, implicating secondary events consistent with breakage–fusion–bridge cycles¹⁶. Spectral karyotyping on three MRR clones showed multiple fusion points confirming rearrangement complexity (Extended Data Table 3). Duplications of chromosome 1 (Fig. 1f, left) and translocations between chromosomes 14 and 11 (Fig. 1f, right) or 14 and 13 were frequently observed from the same clone and even in the same metaphase spread, indicating a role in genome topology¹⁷. Two-colour FISH was performed on a single clone (clone 18 from Extended Data Tables 2 and 3) with the MRR probe and either chromosome 1 or 14. This analysis revealed unstable structures because the MRR could be found at either chromosomes 1 or 14 (Extended Data Fig. 2b), indicating faulty DNA synthesis¹⁸. Furthermore, the MRR pattern changed from a discrete dot to multiple dots interspersed with chromosomal sequences similar to segmental duplications described during evolution¹⁹. Thus, both reporters caused unstable and complex rearrangements, yet the causal pathways are not known.

Complex genomic rearrangements could arise from faulty chromosome maintenance. Therefore, we tested whether γ -radiation or ultraviolet light enhanced repeat fusion for wild-type AB2.2 cells with the IRR or MRR. Exposure to 4 Gy γ -radiation induced repeat fusion for the IRR (Fig. 2a, left, $P = 0.017$, Student's *t*-test) but not the MRR (Fig. 2a, right, $P = 0.16$), whereas exposure to 20 J m⁻² ultraviolet light had the opposite effect on the IRR (Fig. 2b, left, $P = 0.35$) and MRR (Fig. 2b right, $P = 0.006$). This contrast suggests different pathways fused identical and mismatched repeats.

We tested whether homologous recombination proteins fused identical repeats because homologous recombination corrects damage caused by γ -radiation but not ultraviolet light⁴. We tested BLM-defective ES cells (*blm*^{tm3Brd/tm4Brd}, simply called *blm*^{-/-})²⁰ because BLM regulates homologous recombination through Holliday junction dissolution⁵. Repeat fusion was significantly higher in *blm*^{-/-} cells as compared to AB2.2 cells for the IRR (Fig. 2c, compare lanes 1 and 2, $P < 0.0001$), but not the MRR (Fig. 2c, compare lanes 6 and 7, $P = 0.47$). Next we tested *blm*^{-/-} cells haploinsufficient for RAD51 or BRCA2 because BRCA2 enables RAD51 filament formation on DNA single stands to mediate strand annealing and Holliday junction formation. We found *blm*^{-/-} *Rad51*^{+/-} *Aex2*⁻⁴ cells

¹Department of Molecular Medicine/Institute of Biotechnology, The Barshop Institute for Longevity and Aging Studies, The University of Texas Health Science Center at San Antonio, 15355 Lambda Drive, San Antonio, Texas 78245-3207, USA. ²Institute of Molecular Embryology and Genetics (IMEG), Kumamoto University, Honjo 2-2-1 Kumamoto 860-0811, Japan. ³Department of Genetics, Albert Einstein College of Medicine of Yeshiva University, Bronx, New York 10461, USA. [†]Present address: Department of Genetics & Tumor Cell Biology M/S 331, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA.

*These authors contributed equally to this work.

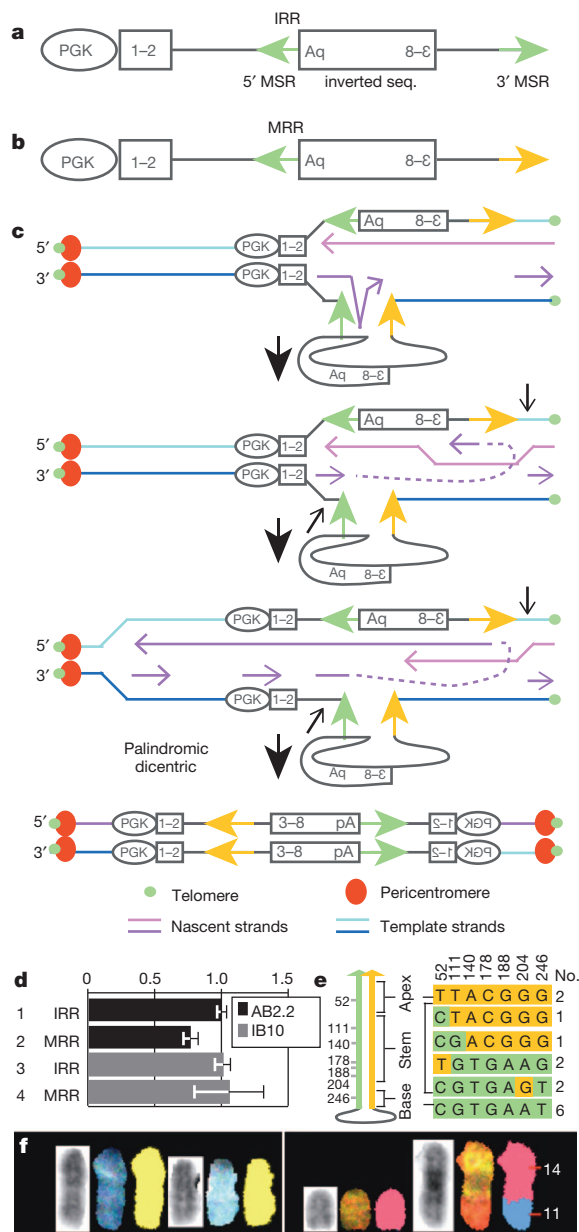


Figure 1 | Inverted repeat fusion. **a, b**, *MiniHPRT* reporters. Promoter (PGK) with intron that separates exons 1 and 2 from 3–8. Repeats at inversion junction. The IRR (**a**) and MRR (**b**) differ only in seven 3' repeat mismatches (green vs orange arrow). **c**, Repeat fusion model. From top to bottom, nascent lagging strand stalls at repeat hairpin and switches to displace complementary template strand to correct *miniHPRT* and produce a dimericentric. **d**, Repeat fusion in AB2.2 and IB10 cells. Shown is the ratio of HAT-resistant colonies compared to IRR. Percentages of HAT-resistant colonies for the IRR in AB2.2 and IB10 are 0.02% and 0.14%, respectively. Biological replicates for lanes 1–4: 19, 19, 18 and 18, respectively. Error bars, standard error of the mean (s.e.m.). **e**, Sequence of fused repeats for the MRR in AB2.2 cells (Extended Data Fig. 1). **f**, Spectral karyotyping analysis on clone 18 (Extended Data Table 3). Left, duplication of chromosome 1; right, translocation of chromosomes 11 and 14.

(Extended Data Fig. 3) and *blm*^{-/-} *brca2*^{+/ Δ ex27-n} cells (Extended Data Fig. 4a) showed reduced repeat fusion (Fig. 2c, compare lane 2 to 3 and 4, $P < 0.0001$). Deleting the remaining *Brca2* exon 27 copy (Extended Data Fig. 4b) further reduced repeat fusion (Fig. 2c, compare lanes 4 and 5, $P = 0.049$). Thus, BLM suppressed RAD51/BRCA2-mediated identical repeat fusion consistent with an homologous-recombination-based pathway (these data do not address the potential role of RAD51/BRCA2 in mismatch repeat fusion).

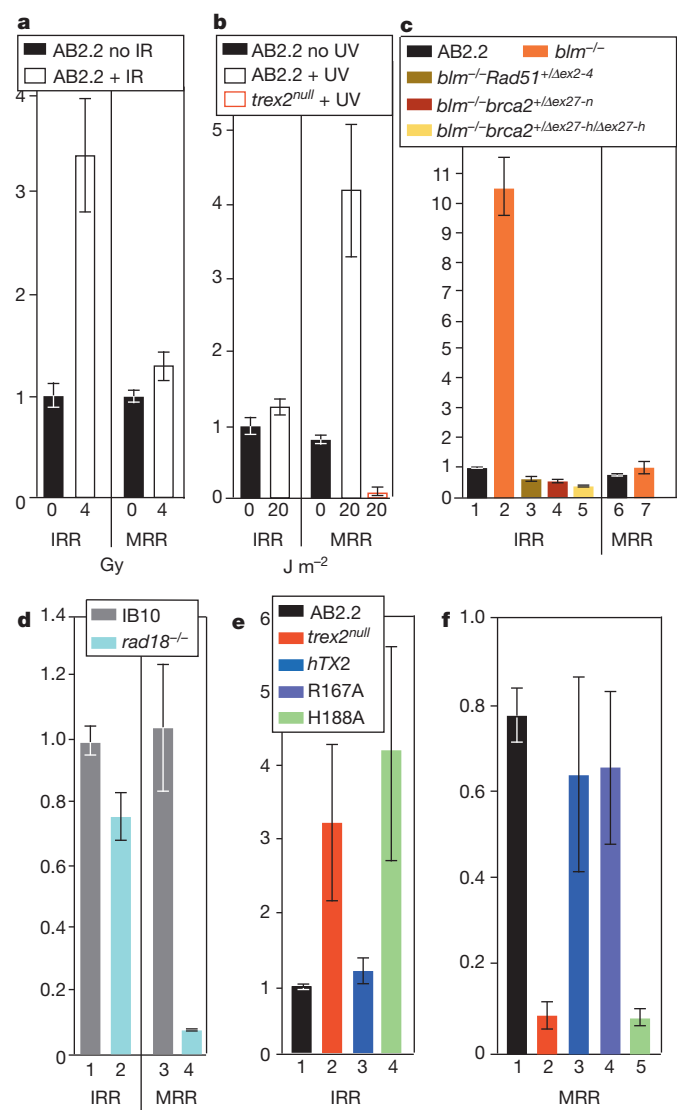


Figure 2 | Two pathways enable repeat fusion that depend on sequence identity. Shown is the ratio of HAT-resistant colonies transfected with IRR in control cells displayed in Fig. 1d. **a**, Gamma radiation (4 Gy) increases fusion for the IRR (left) but not MRR (right). Survival fraction, ~10%. Biological replicates for lanes 1–4: 19, 11, 19 and 11, respectively. **b**, Ultraviolet light (20 J m⁻²) enables fusion for the MRR (right) but not IRR (left). Survival fraction, ~0.6%. Biological replicates for lanes 1–4: 19, 11, 19 and 11, respectively. **c**, BLM suppressed repeat fusion for the IRR but not MRR. *blm*^{-/-} cells deleted for one copy of *Rad51* exons 2–4 (*blm*^{-/-} *Rad51*^{+/ Δ ex2-4} cells), one copy of *Brca2* exon 27 (*blm*^{-/-} *brca2*^{+/ Δ ex27-n}) or two copies of *Brca2* exon 27 (*blm*^{-/-} *brca2*^{+/ Δ ex27-h/ Δ ex27-n}). Biological replicates for lanes 1–7: 19, 23, 12, 12, 12, 19 and 23, respectively. **d**, RAD18 enabled fusion for the MRR more than IRR. Biological replicates: 18 for all lanes. **e, f**, TREX2 suppressed fusion for the IRR (**e**) but enabled fusion for the MRR (**f**). Examined are *trex2*^{null} cells that express human wild-type TREX2 (*hTX2*) or human TREX2 mutated in the DNA binding domain (R167A) or catalytic domain (H188A). Biological replicates for lanes 1–4 in **e**: 19, 19, 20, and 23, respectively, and for lanes 1–5 in **f**: 19, 21, 21, 21 and 23, respectively. Error bars show s.e.m. throughout.

We tested if EF-PRR fused mismatched repeats because ultraviolet light, but not γ -radiation, induced PCNA ubiquitination in mammalian cells²¹. IB10 ES cells deleted for RAD18 (ref. 22) were analysed. These cells showed slightly lower levels of repeat fusion for the IRR as compared to IB10 control cells (Fig. 2d, compare lanes 1 and 2, $P = 0.06$). This reduction could reflect the nonessential participation of RAD18 in homologous recombination¹⁴. By contrast, RAD18-deletion significantly

lowered fusion of mismatched repeats (Fig. 2d, compare lanes 3 and 4, $P = 0.0005$). The reduction of mismatched repeat fusion is greater than identical repeat fusion ($P < 0.0001$), demonstrating that the role of RAD18 in fusing mismatched repeats is more prominent than identical repeats. These results are consistent with EF-PRR fusing mismatched repeats. Yet, RAD18 is an E3 ubiquitin ligase so it could have broad function; therefore, mutations in other genes in the poorly understood EF-PRR pathway should be observed.

TREX2 could be a novel member of EF-PRR. Previously, we analysed *trex2*^{null} cells and cells that expressed wild-type human TREX2 (TREX2^{WT}) and human TREX2 mutated in the catalytic domain (TREX2(H188A)) and DNA-binding domain (TREX2(R167A)), ~85% reduction in DNA binding^{23,24}. We found TREX2 deletion elevated levels of spontaneous isochromatid breaks and chromosomal rearrangements^{24,25}. TREX2^{WT} rescued the null phenotype whereas TREX2(H188A) exacerbated this phenotype, suggesting a dominant effect²⁴. These observations suggested defective DSB repair. However, *trex2*^{null} cells exhibited increased DSB repair and normal BLM-regulated sister chromatid exchanges (SCEs)²⁶. Therefore, we proposed that TREX2 did not repair DSBs but instead suppressed DSB formation through an unknown pathway, possibly EF-PRR. In support, *trex2*^{null} cells had reduced levels of spontaneous SCEs^{26,27}.

TREX2-altered cells were tested for fusion of identical and mismatched repeats. *trex2*^{null} and TREX2(H188A)-expressing cells had elevated levels of identical repeat fusion as compared to control cells (AB2.2 and *Trex2*^{hTX2} cells) (Fig. 2e, compare lanes 1 and 3 to 2 and 4, $P < 0.05$), corroborating our previous observations that homologous recombination is elevated in *trex2*^{null} cells and that an homologous-recombination-based pathway fuses identical repeats. A similar anti-recombination effect on identical repeats was seen for the 3' exonucleases Exo1 and ExoVII in *Escherichia coli*, suggesting that 3' exonuclease activity inhibits these fusions²⁸. We also found *trex2*^{null} and TREX2(H188A)-expressing cells had very low levels of mismatch repeat fusion as compared to AB2.2, *Trex2*^{hTX2} and *Trex2*(R167A) cells (Fig. 2f, compare lanes 1, 3 and 4 to 2 and 5, $P < 0.0006$). Furthermore, TREX2-mediated ultraviolet-light-induced fusion of mismatched repeats (Fig. 2b right panel, $P = 0.003$). These data clearly separate the pathways that mediate identical and mismatch repeat fusion and demonstrate sequence identity determined pathway choice. These data also demonstrate the importance of the catalytic activity of TREX2 in mediating repeat fusion. Exonuclease activity would predictably remove intermediate 3' mismatches or flaps that could occur at the DNA incongruity or during strand exchange and strand displacement. Furthermore, these data are consistent with TREX2 being part of the EF-PRR machinery.

Three experiments were performed to test if TREX2 is a member of EF-PRR. First, TREX2 located to the nascent replication strand after ultraviolet light exposure (Extended Data Fig. 5a); thus, it was at the right place at the right time. Second, TREX2 associated with UBC13, but not MMS2, by glutathione-S-transferase (GST) pull-down (Extended Data Fig. 5b); UBC13–MMS2 is the E2 heterodimer that polyubiquitinates PCNA^{12,21}. In addition, TREX2 associated with UBC13 after ectopic expression in HeLa cells that was enhanced by ultraviolet light (Extended Data Fig. 5c); thus, it associated with the PCNA ubiquitination machinery. Third, we tested the impact TREX2 and RAD18 had on PCNA ubiquitination. As a control we found ultraviolet light, but not γ -radiation, enhanced PCNA ubiquitination as previously seen in human cells²¹ (Extended Data Fig. 6a). TREX2 and RAD18 were needed for efficient PCNA ubiquitination after exposure to ultraviolet light (Extended Data Fig. 6b–d). In addition, cells deleted for both RAD18 and TREX2 (Extended Data Fig. 7) showed no further reduction in PCNA ubiquitination, indicating that they are epistatic (Extended Data Fig. 6b–d). These observations are consistent with TREX2 being part of the EF-PRR machinery and implicate RAD18 and TREX2 in replication fork maintenance.

Potential mechanisms for repeat fusion are faulty DNA repair and faulty DNA replication². Repeat fusion could manifest from faulty

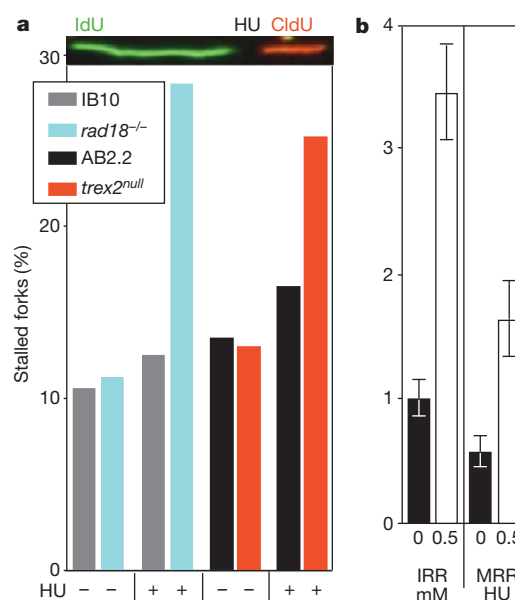


Figure 3 | Hydroxyurea-induced nucleotide depletion. **a**, RAD18 and TREX2 maintain replication forks. The percentage of stalled replication forks after hydroxyurea exposure is shown. Experimental design: cells were cultured in IdU (5-iodo-2'-deoxyuridine) (20 min) to label nascent strand and then exposed to hydroxyurea (0.5 mM, 90 min) to stall replication and then cultured in CldU (5-chloro-2'-deoxyuridine) (20 min) to label restart. Fibre number observed without and with hydroxyurea: IB10 (1,943, 657), *rad18*^{-/-} (1,180, 1,460), AB2.2 (452, 510), *trex2*^{null} (705, 448). **b**, The impact of hydroxyurea (0.5 mM, 90 min) on repeat fusion for the IRR (left) and MRR (right). The ratio of HAT-resistant colonies as compared to AB2.2 cells transfected with the IRR (0.05%) is shown. Survival fraction is 100%. Error bars, s.e.m. Biological replicates, 6 for all lanes.

DNA repair since γ -radiation and ultraviolet light increased fusion. However, the odds that damage actually occurred in or near the reporter sequences is small (even after exposure to agent); thus, the agents could cause a compensatory increase in repair pathways. RAD51, BRCA2 and BLM are involved in both DSB repair and replication fork maintenance^{6,7,10,11,15,29} so either are possible while direct evidence that RAD18 and TREX2 maintain replication forks is lacking in mammalian cells. Therefore, *rad18*^{-/-} and *trex2*^{null} cells were exposed to a brief pulse of low concentration hydroxyurea (0.5 mM, 90 min) that depletes nucleotides to stall replication forks without causing DSBs^{6,7,10,29}. We found *rad18*^{-/-} and *trex2*^{null} cells had elevated levels of stalled replication forks compared to control cells (Fig. 3a, $P < 0.0001$) similar to depletion of the RAD5 orthologue, HLTf³⁰. We further tested faulty replication as causal for repeat fusion by exposing cells with the IRR or MRR to this mild hydroxyurea concentration (Fig. 3b). This exposure increased repeat fusion for the IRR ($P = 0.00025$, Student's *t*-test) and MRR ($P = 0.0037$). Our observations suggest a BLM-regulated pathway consistent with homologous recombination fused identical repeats whereas a RAD18/TREX2-dependent pathway consistent with EF-PRR fused mismatched repeats during replicative stress. These pathways are good candidates for causing complex rearrangements found in cancer and genomic disorders in people and chromosomal variation that leads to species diversification.

METHODS SUMMARY

Repeat fusion assay: the reporters were randomly integrated into ES cells, selected in HAT and colonies counted (colonies were also counted without selection to control for seeding efficiencies). The percentage of HAT-resistant colonies was determined by dividing the number of HAT-resistant colonies by the number of cells electroporated multiplied by the seeding efficiency.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 January; accepted 22 July 2013.

Published online 8 September 2013.

- Hastings, P. J., Lupski, J. R., Rosenberg, S. M. & Ira, G. Mechanisms of change in gene copy number. *Nature Rev. Genet.* **10**, 551–564 (2009).
- Carr, A. M., Paek, A. L. & Weinert, T. DNA replication: failures and inverted fusions. *Semin. Cell Dev. Biol.* **22**, 866–874 (2011).
- Lee, J. A., Carvalho, C. M. & Lupski, J. R. A. DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131**, 1235–1247 (2007).
- San Filippo, J., Sung, P. & Klein, H. Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem.* **77**, 229–257 (2008).
- Wu, L. & Hickson, I. D. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* **426**, 870–874 (2003).
- Schlacher, K. *et al.* Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. *Cell* **145**, 529–542 (2011).
- Schlacher, K., Wu, H. & Jasin, M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell* **22**, 106–116 (2012).
- Mizuno, K., Lambert, S., Baldacci, G., Murray, J. M. & Carr, A. M. Nearby inverted repeats fuse to generate acentric and dicentric palindromic chromosomes by a replication template exchange mechanism. *Genes Dev.* **23**, 2876–2886 (2009).
- Mizuno, K., Miyabe, I., Schalbetter, S. A., Carr, A. M. & Murray, J. M. Recombination-restarted replication makes inverted chromosome fusions at inverted repeats. *Nature* **493**, 246–249 (2013).
- Petermann, E., Orta, M. L., Issaeva, N., Schultz, N. & Helleday, T. Hydroxyurea-stalled replication forks become progressively inactivated and require two different RAD51-mediated pathways for restart and repair. *Mol. Cell* **37**, 492–502 (2010).
- Carr, A. M. & Lambert, S. Replication stress-induced genome instability: the dark side of replication maintenance by homologous recombination. *J. Mol. Biol.* <http://dx.doi.org/10.1016/j.jmb.2013.04.023> (30 April 2013).
- Ulrich, H. D. Regulating post-translational modifications of the eukaryotic replication clamp PCNA. *DNA Repair (Amst.)* **8**, 461–469 (2009).
- Falbo, K. B. *et al.* Involvement of a chromatin remodeling complex in damage tolerance during DNA replication. *Nature Struct. Mol. Biol.* **16**, 1167–1172 (2009).
- Huang, J. *et al.* RAD18 transmits DNA damage signalling to elicit homologous recombination repair. *Nature Cell Biol.* **11**, 592–603 (2009).
- Kim, T. M. *et al.* RAD51 mutants cause replication defects and chromosomal instability. *Mol. Cell Biol.* **32**, 3663–3680 (2012).
- Shimizu, N., Shingaki, K., Kaneko-Sasaguri, Y., Hashizume, T. & Kanda, T. When, where and how the bridge breaks: anaphase bridge breakage plays a crucial role in gene amplification and HSR generation. *Exp. Cell Res.* **302**, 233–243 (2005).
- Cavalli, G. & Misteli, T. Functional implications of genome topology. *Nature Struct. Mol. Biol.* **20**, 290–299 (2013).
- Harada, S., Sekiguchi, N. & Shimizu, N. Amplification of a plasmid bearing a mammalian replication initiation region in chromosomal and extrachromosomal contexts. *Nucleic Acids Res.* **39**, 958–969 (2011).
- Horvath, J. E. *et al.* Using a pericentromeric interspersed repeat to recapitulate the phylogeny and expansion of human centromeric segmental duplications. *Mol. Biol. Evol.* **20**, 1463–1479 (2003).
- Luo, G. *et al.* Cancer predisposition caused by elevated mitotic recombination in Bloom mice. *Nature Genet.* **26**, 424–429 (2000).
- Motegi, A. *et al.* Polyubiquitination of proliferating cell nuclear antigen by HLTf and SHPRH prevents genomic instability from stalled replication forks. *Proc. Natl Acad. Sci. USA* **105**, 12411–12416 (2008).
- Tateishi, S. *et al.* Enhanced genomic instability and defective postreplication repair in RAD18 knockout mouse embryonic stem cells. *Mol. Cell Biol.* **23**, 474–481 (2003).
- Chen, M. J., Ma, S. M., Dumitrache, L. C. & Hasty, P. Biochemical and cellular characteristics of the 3' → 5' exonuclease TREX2. *Nucleic Acids Res.* **35**, 2682–2694 (2007).
- Dumitrache, L. C., Hu, L. & Hasty, P. TREX2 exonuclease defective cells exhibit double-strand breaks and chromosomal fragments but not Robertsonian translocations. *Mutat. Res.* **662**, 84–87 (2009).
- Chen, M. J. *et al.* Cisplatin depletes TREX2 and causes Robertsonian translocations as seen in TREX2 knockout cells. *Cancer Res.* **67**, 9077–9083 (2007).
- Dumitrache, L. C. *et al.* Trex2 enables spontaneous sister chromatid exchanges without facilitating DNA double-strand break repair. *Genetics* **188**, 787–797 (2011).
- Goldfless, S. J., Morag, A. S., Belisle, K. A., Suter, V. A. Jr & Lovett, S. T. DNA repeat rearrangements mediated by DnaK-dependent replication fork repair. *Mol. Cell* **21**, 595–604 (2006).
- Dutra, B. E. & Lovett, S. T. Cis and trans-acting effects on a mutational hotspot involving a replication template switch. *J. Mol. Biol.* **356**, 300–311 (2006).
- Sirbu, B. M. *et al.* Analysis of protein dynamics at active, stalled, and collapsed replication forks. *Genes Dev.* **25**, 1320–1327 (2011).
- Blastyák, A., Hajdú, I., Unk, I. & Haracska, L. Role of double-stranded DNA translocase activity of human HLTf in replication of damaged DNA. *Mol. Cell Biol.* **30**, 684–693 (2010).

Acknowledgements We thank C. Williams and S. Dodds for technical assistance and the Molecular Cytogenetic Core at Albert Einstein College of Medicine for help with the execution of spectral karyotyping and two-colour FISH. This work was supported by the National Institutes of Health (1 RO1 CA123203-01A1 to P.H., 2P01AG017242-12 to P.H. and C.M. P30CA013330 to C.M.) and with support from the Cancer Therapy & Research Center at The University of Texas at San Antonio (CTRC) (P30 CA054174).

Author Contributions L.H., T.M.K., P.R.Y., C.M., L.C.D. and P.H. designed experiments and interpreted results. L.H., T.M.K., M.Y.S., S.-AK., C.L.H., D.H.K. and P.H. performed experiments. S.T. provided the *rad18*^{-/-} and IB10 ES cells. P.H. wrote the paper with comments from the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.H. (hasty@uthscsa.edu).

METHODS

Construction of the IRR and MRR. The IRR and MRR contain a puromycin phosphotransferase (puro) selection cassette and an *HPRT* minigene³¹ (*miniHPRT*). Puro was positioned 5' to *miniHPRT* and used to select for stable transfectants. *MiniHPRT* contains a phosphoglycerate kinase 1 (PGK) promoter³², exons 1 and 2, intron and exons 3–8 with polyadenylation sequences. The 3' half of *miniHPRT* was inverted from intronic Xba1. Major satellite repeats (MSRs)³³ were positioned at inversion junctions in an indirect orientation. The same MSR sequence (below) is located at both junctions for the IRR (Fig. 1a, green arrow) and at the 5' junction for the MRR (Fig. 1b, green arrow) while a divergent MSR (seven mismatches) is located at the 3' end for the MRR (Fig. 1b, orange arrow). These mismatches are the only difference between the reporters.

MSR sequence, mismatched nucleotides underlined (Fig. 1a, b, green arrow):

5'-TGGAATATGGCGAGAAAACTGAAATCATGGAAATGAGAAATA
CACACTTCAGGACGTGAAATATGGCGAGGAAAACTGAAAAAGGTGGAA
AATTGAGAAATGTCCACTGTAGGACGTGGAATATGGCAAGAAAACTGT-
AAATCATGGAAATGAGAAACATCCACTTGACGACTTGAAAAATGACA-
AAATCACTAAAAAACATGAAAAATGAGAAATGCACACTGAAGGACCTG
GAATATGGCTAGAAAACTGAAATCACGGAAATGAGAAATACAAA
CCTTAGGACTTGAAATATGGCGAGGAAAACT3'

MSR sequence, mismatched nucleotides are underlined (Fig. 1b, orange arrow)

5'-TGGAATATGGCGAGAAAACTGAAATCATGGAAATGAGAAATA
CACACTTCAGGACGTGAAATATGGCGAGGAAAACTGAAAAAGGTGGAA
AATTGAGAAATGTCCACTGTAGGACGTGGAATATGGCAAGAAAACT-
GAAATCATGGAAATGAGAAACATCCACTTGACGACTTGAAAAATG-
ACGAAATCACTAAAAAACGTGAAAAATGAGAAATGCACACTGAAGGA
CCTGGAATATGGCGAGAAAACTGAAATCACGGAAATGAGAAATAC
AAACCTTAGGACTTGAAATATGGCGAGGAAAACTG-3'

PCR amplification of repeat fusion. PCR amplify fusions with primers 5' (HPRT4) and 3' (HPRT recom Rev) to Xba1. Sequence PCR products with the same primers.

HPRT4, 5'-TCTCAAGCACTGGCCTATGC 3'; HPRT recom Rev, 5'-AGACAGAATGCTATGCAACC-3'.

Conditions: 1 cycle at 98 °C for 10 min, 35 cycles: 98 °C for 1 min, 62 °C for 1 min, 72 °C for 20 s.

Tissue culture for mouse ES cells. Maintain ES cells in M15 (high glucose DMEM with 15% fetal bovine serum, 100 µM β-mercaptoethanol, 2 mM L-glutamine, 3 mg ml⁻¹ penicillin, 5 mg ml⁻¹ streptomycin, 1,000 U ml⁻¹ ESGRO (LIF)) on plastic plates precoated with gelatin (0.1%, ~1 h) and seeded with 2.5 × 10⁶ primary murine embryonic fibroblasts (MEFs, mutated for *Hprt* and resistant to puromycin, exposed to 30 Gy γ-irradiation) and incubate at atmospheric O₂, 5% CO₂, 37 °C. ES cells were also cultured on gelatinized plates without feeders.

Repeat fusion assay. Repeat fusion is seen in cells transfected with the IRR or MRR (Figs 1d, 2 and 3b). Transfect ES cells (5 × 10⁶, 800 µl PBS) with 5 µg of uncut IRR or MRR by electroporation (Bio-Rad Gene Pulsar at 230 V, 500 µF). Seed cells onto 3–6 3.5-cm plates with mitotically inactivated MEFs. Each well is a replicate because they remain separate. Add puromycin (3 µg ml⁻¹) next day. About 100–200 puromycin resistant colonies grow for each well. Seven days later, pool puromycin-resistant colonies for each well and passage onto a 3.5-cm plate precoated with gelatin. Three days later passage cells onto a 10-cm plate precoated with gelatin. See below for cell exposure to DNA-damaging agents. For unexposed cells, next day seed 10⁶ cells onto a gelatin-coated 10-cm plastic plate in M15 supplemented with 1 × HAT (1 mM sodium hypoxanthine, 4 µM aminopterin and 160 µM thymidine). Count HAT-resistant colonies 10 days later. To control for seeding efficiencies, seed 2,000 cells for each replicate onto a gelatin-coated 3.5-cm plastic plate and culture in M15 without selection. Determine percentage of HAT-resistant colonies by dividing the number of HAT-resistant colonies by the number of cells electroporated multiplied by the seeding efficiency.

For cells exposed to DNA-damaging agents (γ-radiation or ultraviolet light or hydroxyurea (HU)) the protocol is the same for the transfection, selection in puromycin and expansion of puromycin resistant cells (see earlier). After expansion, expose cells to either 4 Gy γ-radiation (¹³⁷Cs at a rate of 0.125 Gy s⁻¹, Mark1 gamma radiation source from Shepard and Associates) or 20 J m⁻² ultraviolet light (a dual wavelength ultraviolet transilluminator from Alpha Innotech at a rate of 1 J m⁻² s⁻¹) or HU (0.5 mM, 90 min). For γ-radiation and ultraviolet light, expose cells directly on the plate after removing media. Then add 10 ml of pre-warmed (37 °C) fresh media and incubated for 48 h. Then seed 10⁶ cells onto a gelatin-coated 10-cm plastic plate in M15 supplemented with 1 × HAT. Count HAT-resistant colonies 10 days later. To control for seeding efficiencies and survival fraction, seed 2,000 cells for each replicate onto a gelatin-coated 3.5-cm plastic plate and culture in M15 without selection. Survival fraction is ~10%, 0.6% or 100% after exposure to γ-radiation (4 Gy), ultraviolet (20 J m⁻²) or HU (0.5 mM, 90 min), respectively.

Two-colour FISH with the pericentromeric and telomeric probes. Perform two-colour FISH (Extended Data Fig. 2a) on HAT-resistant colonies expanded with the IRR or MRR. Seed cells in HAT selection media on plastic plates pre-coated with gelatin. Next day add fresh media (without HAT). Treat cells with colcemid (540 nM, 4 h) then trypsinize. Slide preparation: spin cells (180g, for 10 min), wash twice in PBS (pH 7.4) and resuspend pellet in 300 µl 75 mM KCl, drop-wise, flicking tube. Incubate in a 37 °C water bath (15 min). Add drop-wise 300 µl methanol/acetic acid (2:1 fixative) while flicking tube, spin 845g, 30 s. Wash cells in 300 µl 2:1 fixative, drop-wise, flicking tube, spin 850g, 30 s; repeat wash. Hybridization: place slides in methanol overnight, then incubate in 70% formamide at 70 °C, place slides in 30% formamide at 37 °C in dark with 500 µl per slide of 0.25 mg ml⁻¹ pericentromeric (CY-3 5'-TGGAATATGGCGAGAAAACTGAAATCATGGAAATGAGAA-3') and telomeric (6-FAM 5'-(CCCTAA)₇ 3') probes for 15 min, wash in PBS, 10 dips, coverslip in 4',6-diamidino-2-phenylindole (DAPI).

Spectral karyotyping. Perform spectral karyotyping (Fig. 1f) as described³⁴ with commercial spectral karyotyping paint probes from Applied Spectral Imaging. Define rearrangements with nomenclature rules from the International Committee on Standard Genetic Nomenclature for Mice³⁵.

Two-colour FISH with the MRR and chromosome 1 or 14 paint. Perform two-colour FISH (Extended Data Fig. 2b) with custom-made chromosome paint probes specific for murine autosomes 1 and 14 labelled with the Spectrum Green (Dyomics) using a standard degenerate oligonucleotide primed-polymerase chain reaction (DOP-PCR) protocol (<http://atlasgeneticsoncology.org/Deep/ComparCancerCytogID20011.html>). Label MRR with Spectrum Orange dUTP (Dyomics) by nick-translation and hybridize to chromosomal preparations derived from clone 18 (Extended Data Table 3). After overnight hybridization (37 °C), wash slides and counterstain with DAPI and image random fields with an inverted Zeiss Axiovert 200 using fine focusing oil immersion lens (×60, numerical aperture 1.35). Equip microscope with a Camera Hall 100 and Applied Spectral Imaging software.

Generation of mouse *Rad51* targeting vector. Construct mouse *Rad51* targeting vector (Extended Data Fig. 3) as described³⁶. Amplify left (5') and right (3') homologous arms with high-fidelity PCR using genomic DNA extracted from AB2.2 ES cells and iProof DNA polymerase (Bio-Rad Laboratories) in 25 µl containing 5 µl of 5 × iProof HF buffer, 0.5 µl of 10 mmol l⁻¹ deoxynucleotide triphosphates, 0.75 µl of 4 µmol l⁻¹ forward and reverse primers (below), 100 ng of genomic DNA, and 0.25 µl of iProof DNA polymerase.

Left arm primers: *Rad51*KiLa forward, 5'-CACACTCGAGTCCCTCTA CGCTGAGAAGCCGGAGAAAG-3'; *Rad51*KiLa reverse, 5'-CACAGCGGCCG CAGGCCACTAAGGCCAGAACTGCAGCTGGCCCTCCCTATCCAC-3'.

Right arm primers: *Rad51*KiRa forward, 5'-CACAGCGGCCGAGGCCT GCGTGGCCGAGTATAGGAATGTTCAGCTTCTCATAGAC-3'; *Rad51*KiRa reverse, 5'-CACAGTCGACGGTACTGGTTAGTTCATAATGTTGTTCCA-3'.

PCR conditions for both arms: 1 cycle: 98 °C for 5 min 35 cycles: 98 °C for 1 min, 64.7–70.2 °C gradient for 1 min, 72 °C for 1 min and 30 s. 1 cycle: 72 °C for 10 min.

After amplifying arms, digest left arm (3.9 kb) with Sall and NotI and clone into a plasmid backbone, pKO, cut with XhoI and NotI. Then, digest right arm (3.0 kb) with XhoI and NotI and clone into the same backbone digested with Sall and NotI to delete *Rad51* exons 2–4. Then, clone floxed SAβgeo-*miniHPRT* (Extended Data Fig. 3a) into unique SfiI sites as described³⁶.

Transfect targeting vector (5 µg, cut with PacI) into *blm*^{-/-} ES cells (5 × 10⁶ cells in 800 µl PBS) by electroporation (Gene Pulser Cuvettes with a 0.4 cm electrode gap at 230 V, 500 µF with a Gene Pulser Apparatus from Bio-Rad). After electroporation, seed cells onto two 10 cm plates with mitotically inactive MEFs. Next day, add M15 medium containing 1 × HAT (0.1 mM hypoxanthine, 0.0004 mM aminopterin and 0.016 mM thymidine). Pick HAT-resistant colonies 7 days later onto a 96-well plate and maintain in HAT selection. Replica plate to freeze one plate and use the other to isolate genomic DNA³⁷. Screen for targeted clones with PCR (Extended Data Fig. 3b).

H13F (in *miniHPRT*): 5'-GTAAATGAAAAATTCTCTTAAACCACAGCA CTATTGAG-3' SR3 (outside the right arm): 5'-AGCCAGGTATAGTCTCAAA GGAATCTGCAATCC-3'.

PCR conditions: 1 cycle: 98 °C for 5 min; 35 cycles: 98 °C for 1 min, 67 °C for 1 min, 72 °C for 1 min 30 s; and 1 cycle: 72 °C for 10 min.

Cre-mediated deletion of SAβgeo and 5' *miniHPRT*. Delete SAβgeo and 5' half of *miniHPRT* using Cre recombinase to generate *Rad51*^{+/Δex2-4} cells (Extended Data Fig. 3c). Expand targeted ES cells in 1 × HAT to remove HPRT-negative cells that survive due to cross feeding. Removed HAT selection 2 days before transfection and cultured in 1 × HT (1 mM sodium hypoxanthine and 160 µM thymidine); electroporate 5 × 10⁶ cells in 800 µl DPBS with 10 µg of pPGKcreA using a Bio-Rad Gene Pulsar at 230 V, 500 µF. After electroporation, seed 200 µl onto a 10-cm feeder plate without selection for 2–4 days to allow time for *miniHPRT* removal

and time for degradation of *HPRT* mRNA and protein. Then seed 4×10^4 cells onto a 10-cm feeder plate in $10 \mu\text{M}$ 6-thioguanine. Pick 6-thioguanine-resistant colonies 10 days later. Expand cells in $10 \mu\text{M}$ 6-thioguanine and replica plate. Freeze one plate and use the other to isolate genomic DNA³⁷. Confirmed Cre-mediated deletion with PCR (1.4 kb fragment).

PCR primers. RCF1 (in *RAD51* intron 1), 5'-GTGCTGAATCTCTAGAA CTG-3'; AS2 (in exon cluster 3-8 of *miniHPRT*), 5'-TGTCCTGTGACTGG TCA-3'.

PCR conditions: 1 cycle: 98°C for 5 min, 35 cycles: 98°C for 1 min, 64°C for 1 min, 72°C for 30 s. 1 cycle: 72°C for 10 min.

Targeting mouse *Brca2* exon 27. Replace the first copy of *Brca2* exon 27 with PGKneobpA³⁸ by cloning PGKneobpA into the Sfi1 sites of the *Brca2* exon 27 deletion targeting vector (Extended Data Fig. 4a)³⁶. Transfect as described for *Rad51*. Use PCR to detect targeted clones (Extended Data Fig. 4a).

PCR primers: NF (in neo), 5'-AGCGCATCGCCTTCTATCGCCTTCTTG ACG-3'; *Brca2* intron 27 reverse, 5'-CCCCGTCGACCGGAGAGCTAATGG CCTTACTCCAACG-3'. Conditions: 35 cycles of 98°C for 1 min, 65°C for 1 min, 72°C for 1 min and 30 s.

Replace the second copy of *Brca2* exon 27 with floxed *miniHPRT* (Extended Data Fig. 4b)³⁶. Use PCR to detect targeted clones (Extended Data Fig. 4b).

PCR primers: H13F, 5'-GTAAATGAAAAAATTCTCTTAACACAGC ACTATTGAG-3'; B27R, 5'-CCCCGTCGACCGGAGAGCTAATGGCCTCTAC TCCAACG-3'. Conditions: 35 cycles of 98°C for 1 min, 65°C for 1 min, 72°C for 1 min and 30 s.

Removed the 5' half of *miniHPRT* by Cre-mediated recombination³⁶ to generate *Brca2*^{Δex27-hiAex27-n} cells. Use PCR to detect removal (Extended Data Fig. 4b).

PCR primers: Bi26, 5'-TCAATCAAGCAGTCCTCACC-3'; H3-8R: 5'-TGACAGTCAACAGGGGACA-3'. Conditions: 35 cycles: 98°C for 1 min, 65°C for 1 min, 72°C for 45 s.

Coimmunoprecipitation of IdU and Myc-TREX2 after exposure to ultraviolet light. TREX2 associates with nascent strand DNA after ultraviolet exposure (Extended Data Fig. 5a). Experiment performed as described¹⁰ with minor modifications. Transfected HeLa cells with $5 \mu\text{g}$ Myc-TREX2 using FuGENE6 (Roche). Label cells with IdU ($5 \mu\text{M}$, 30 min), treat with 20 J m^{-2} ultraviolet and recover with the indicated time. Crosslink cells in formaldehyde (1%, 15 min 24°C). Remove cytoplasmic protein fraction by incubation in hypotonic buffer (10 mM HEPES, 1.5 mM MgCl_2 , 10 mM KCl, β -mercaptoethanol, PMSF, Protease Inhibitor (Roche) for 10 min on ice). Resuspend pellets in nuclear exact buffer (20 mM HEPES, 20% glycerol, 400 mM KCl, 1.5 mM MgCl_2 , 0.2 mM EDTA, β -mercaptoethanol, PMSF, Protease Inhibitor (Roche)). Dilute nuclear exact protein ($50 \mu\text{g}$) solution with equal volume of immunoprecipitation dilution buffer (20 mM HEPES, 0.2 mM EDTA, 10% glycerol, PMSF, Protease Inhibitor (Roche)) and pre-wash with Protein G Sepharose beads ($10 \mu\text{l}$, 1 h). Remove bead and immunoprecipitate supernatant by incubating with $1 \mu\text{g}$ of anti-BrdU (mouse anti-BrdU B44) at 4°C overnight. Incubate reaction solution with $20 \mu\text{l}$ Protein G Sepharose beads for 3 h at 4°C and wash beads 4 times with immunoprecipitation wash buffer. Separate immunoprecipitated proteins with SDS-PAGE gel and blot with anti-Myc (BD Bioscience) antibody.

TREX2-UBC13 association. TREX2 associates with UBC13 by GST pull-down (Extended Data Fig. 5b). Bind GST-MMS2, GST-UBC13, and GST-TREX2 fusion proteins ($5 \mu\text{g}$) to glutathione-Sepharose 4B (GE Healthcare) and incubate with [³⁵S]methionine-labelled TREX2 ($4 \mu\text{l}$, 1.5 h, 23°C)³⁹. Wash beads with NETN buffer (50 mM Tris, 250 mM NaCl, 5 mM EDTA, pH 7.5, and 0.1% NP40) and subject to SDS-PAGE and phosphorimager analysis.

TREX2 associates with UBC13 by coimmunoprecipitation in HeLa cells (Extended Data Fig. 5c). Transfect HeLa cells with $5 \mu\text{g}$ Myc-TREX2 and $5 \mu\text{g}$ HA-UBC13 plasmid (48 h) using FuGENE6 (Roche), expose cells to 0 J m^{-2} or 20 J m^{-2} ultraviolet as described for the PCNA ubiquitination assay (below). Crosslink cells in formaldehyde (1%, 15 min, 24°C). Incubate in hypotonic buffer (10 mM HEPES, 1.5 mM MgCl_2 , 10 mM KCl, β -mercaptoethanol, PMSF, Protease Inhibitor (Roche)) for 10 min on ice to remove cytoplasmic protein fraction. Resuspend pellets in nuclear exact buffer (20 mM HEPES, 20% glycerol, 400 mM KCl, 1.5 mM MgCl_2 , 0.2 mM EDTA, β -mercaptoethanol, PMSF, Protease Inhibitor (Roche)). Dilute nuclear exact protein ($50 \mu\text{g}$) solution with equal volumes of immunoprecipitation dilution buffer (20 mM HEPES, 0.2 mM EDTA, 10% glycerol, PMSF, Protease Inhibitor (Roche)) and incubate with $10 \mu\text{l}$ Protein G Sepharose beads (1 h). Remove beads and immunoprecipitate supernatant by incubating with $2 \mu\text{g}$ anti-Myc (BD Bioscience) or anti-HA (Roche) antibody at 4°C overnight. Incubate reaction solution with $20 \mu\text{l}$ Protein G Sepharose beads for 3 h. Wash beads 4 times with immunoprecipitation wash buffer. Separate immunoprecipitated proteins by SDS-PAGE gel and blot with anti-Myc or anti-HA antibody.

Detection of PCNA ubiquitination with chromatin-bound fraction. RAD18 and TREX2 participated in ultraviolet-induced PCNA ubiquitination (Extended

Data Fig. 6). Isolate chromatin-bound fraction as described^{21,40} with modifications. Briefly, resuspend $\sim 1.5 \times 10^7$ cells in buffer A (10 mM HEPES (pH 7.9), 1.5 mM MgCl_2 , 10 mM KCl, 0.34 M sucrose, 10% glycerol, 0.1% Triton X-100 and protease inhibitor cocktail (Roche)), incubate and rotate 5 min at 4°C then centrifuge (5,204g, 2 min, 4°C). Remove soluble fraction. Resuspended pellet in buffer then centrifuge (5,204g, 3 min, 4°C). Extract chromatin-bound fraction, resuspend pellet in buffer B (20 mM Tris-Cl (pH 8.1), 2 mM EDTA (pH 8.0), 500 mM NaCl, 0.1% SDS, 1% Triton X-100 and protease inhibitor cocktail (Roche)), sonicate, treat with micrococcal nuclease (10 min, 37°C) and centrifuge (17,948g, 15 min, 4°C). Immunoprecipitate supernatant containing released chromatin-bound protein. Pre-incubate with protein G Sepharose beads (GE healthcare) (1–2 h, 4°C) to pre-cleaned protein and incubated with $1 \mu\text{g}$ of anti-PCNA antibody (PC10, Santa Cruz Biotechnology) overnight at 4°C . Precipitate anti-PCNA immune complexes with $30 \mu\text{l}$ protein G Sepharose beads for 3 h at 4°C . Separate protein on 10% SDS-PAGE gel and transfer onto PVDF membrane. Use monoclonal antibodies for western blot: anti-ubiquitin (P4D1, COVANCE; 1:1,000–2,000) or anti-PCNA (PC10, Santa Cruz Biotechnology; 1:2,000–2,500). Used mouse TrueBlot ULTRA (Anti-mouse Ig HRP, ROCKLAND; 1:1,000–2,500) to minimize IgG signal. Quantify band intensities with ImageJ software (<http://rsbweb.nih.gov/ij/>).

Targeting *Trex2* in IB10 cells and *rad18*^{-/-} cells. Electroporate *Trex2* targeting vector ($5 \mu\text{g}$ of PacI-cut) (Extended Data Fig. 7)²⁵ into IB10 cells and *rad18*^{-/-} cells as described for *Rad51*.

Primers to detect left arm integration: TX2 LR55 (outside of left arm), 5'-TATATTTAGGAGACAAAGTGGCCCTGCCAGAGCTG-3'; HATrev (in the *HPRT* minigene), 5'-CATGCGCTTATAGCAGCCCGCTGGGCACTTGGCGC-3'. Conditions: 1 cycle: 98°C for 5 min 35 cycles: 98°C for 1 min, 72°C for 1 min, 72°C for 2 min 30 s. 1 cycle: 72°C for 10 min.

Primers to detect right arm integration: HATfor (in the *HPRT* minigene), 5'-GTAAATGAAAAAATTCTCTTAACACACAGCACTATTGAG-3'; TX2 RR33 (outside the right arm), 5'-CCTGTTTCACAAATATCAGGACCTGAGTTT GTATCC-3'. Conditions: 1 cycle: 98°C for 5 min 35 cycles: 98°C for 1 min, 63.5°C for 1 min, 72°C for 2 min 30 s. 1 cycle: 72°C for 10 min.

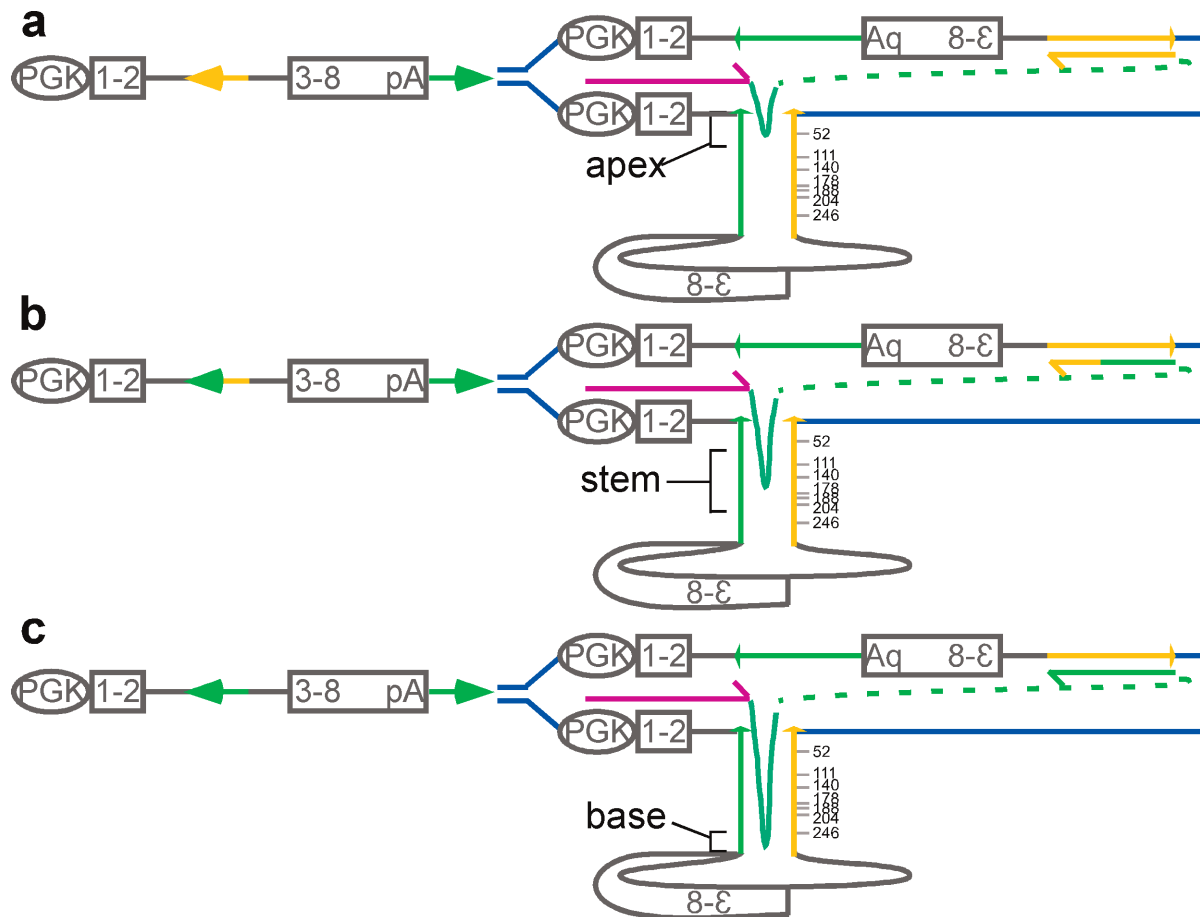
Primers to confirm deletion of TREX2 open reading frame: mTX2For, 5'-AAAAGAATTCGCCACCATGTCTGAGCCACCCCGGGC-3'; mTX2Rev: 5'-AAACTCGAGTCAGGCTTCGAGGCTTGGACC-3'. Conditions: 1 cycle: 98°C for 5 min 35 cycles: of 98°C for 1 min, 65°C for 1 min, 72°C for 25 min. 1 cycle: 72°C for 10 min.

Microfibre analysis. RAD18 and TREX2 enabled replication fork restart (Fig. 3a). Perform DNA fibre analysis as described^{10,15} with modifications. Pulse-label ES cells with IdU ($25 \mu\text{M}$, 20 min), wash twice with medium, expose to HU (0.5 mM, 1.5 h), wash twice with medium and pulse-label with CldU ($250 \mu\text{M}$, 20 min). Fix fibres in methanol and acetic acid (3:1) and air-dry. To denature fibres, treat slides with HCl (2.5 M, 75–80 min) and wash twice with PBS then block 1 h with 1% BSA (bovine serum albumin) + 0.1% Tween 20. Incubate slides with primary antibodies against CldU (rat anti-BrdU BU1/75[ICRI], Abcam, 1:1,000) and IdU (mouse anti-BrdU B44, 1:750) for 1.5 h. Fix slides with 4% paraformaldehyde and wash thrice with PBS. Apply AlexaFluor 555-conjugated goat anti-rat IgG (Molecular Probes, 1:500) and AlexaFluor 488-conjugated goat anti-mouse IgG (Molecular Probes, 1:500) to slides for 2 h. Wash slides and mount in Fluoroshield (Sigma) and examine (Axioplan2, Zeiss fluorescent microscope).

Statistics. Student's *t*-test was used for statistics (two-sided without adjustments for multiple comparisons). The average was the centre value. In all figures the s.e.m. is shown and the number of biological replicates are provided in the legends.

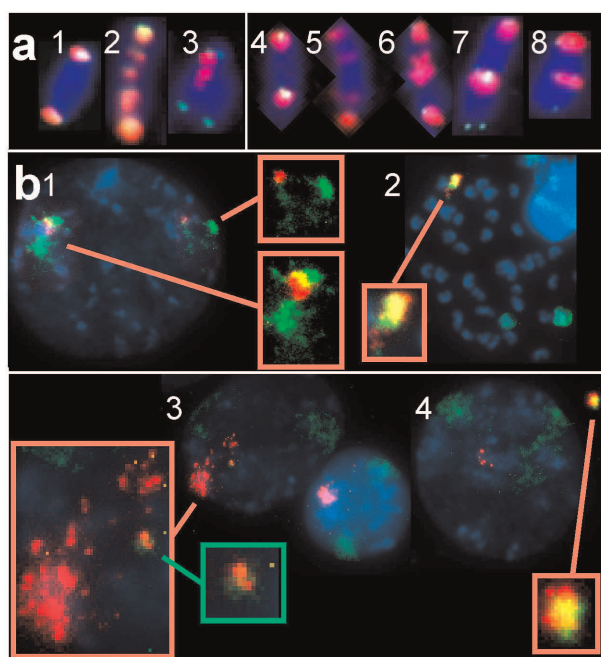
- Reid, L. H., Gregg, R. G., Smithies, O. & Koller, B. H. Regulatory elements in the introns of the human *HPRT* gene are necessary for its expression in embryonic stem cells. *Proc. Natl Acad. Sci. USA* **87**, 4299–4303 (1990).
- Adra, C. N., Boer, P. H. & McBurney, M. W. Cloning and expression of the mouse *pgk-1* gene and the nucleotide sequence of its promoter. *Gene* **60**, 65–74 (1987).
- Guenatri, M., Bailly, D., Maison, C. & Almouzni, G. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J. Cell Biol.* **166**, 493–505 (2004).
- Montagna, C., Andrechek, E. R., Padilla-Nash, H., Muller, W. J. & Ried, T. Centrosome abnormalities, recurring deletions of chromosome 4, and genomic amplification of *HER2/neu* define mouse mammary gland adenocarcinomas induced by mutant *HER2/neu*. *Oncogene* **21**, 890–898 (2002).
- Davison, M. T. Rules and guidelines for nomenclature of mouse genes. *Gene* **147**, 157–160 (1994).
- Holcomb, V. B. *et al.* *HPRT* minigene generates chimeric transcripts as a by-product of gene targeting. *Genesis* **45**, 275–281 (2007).
- Ramírez-Solis, R. *et al.* Genomic DNA microextraction: a method to screen numerous samples. *Anal. Biochem.* **201**, 331–335 (1992).
- Soriano, P., Montgomery, C., Geske, R. & Bradley, A. Targeted disruption of the *c-src* proto-oncogene leads to osteopetrosis in mice. *Cell* **64**, 693–702 (1991).
- Kim, D. H. *et al.* The CRL4^{Cdt2} ubiquitin ligase mediates the proteolysis of cyclin-dependent kinase inhibitor Xic1 through a direct association with PCNA. *Mol. Cell Biol.* **30**, 4120–4133 (2010).

40. Krijger, P. H. *et al.* HLTf and SHPRH are not essential for PCNA polyubiquitination, survival and somatic hypermutation: existence of an alternative E3 ligase. *DNA Repair (Amst.)* **10**, 438–444 (2011).
41. Friedrich, G. & Soriano, P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* **5**, 1513–1523 (1991).
42. Araki, K., Araki, M. & Yamamura, K. Targeted integration of DNA using mutant lox sites in embryonic stem cells. *Nucleic Acids Res.* **25**, 868–872 (1997).
43. Kim, T. M., Choi, Y. J., Ko, J. H. & Hasty, P. High-throughput knock-in coupling gene targeting with the *HPRT* minigene and Cre-mediated recombination. *Genesis* **46**, 732–737 (2008).
44. Donoho, G. *et al.* Deletion of *Brca2* exon 27 causes hypersensitivity to DNA crosslinks, chromosomal instability, and reduced life span in mice. *Genes Chromosom. Cancer* **36**, 317–331 (2003).
45. Morimatsu, M., Donoho, G. & Hasty, P. Cells deleted for *Brca2* COOH terminus exhibit hypersensitivity to gamma- radiation and premature senescence. *Cancer Res.* **58**, 3441–3447 (1998).
46. Moynahan, M. E., Pierce, A. J. & Jasin, M. BRCA2 is required for homology-directed repair of chromosomal breaks. *Mol. Cell* **7**, 263–272 (2001).
47. Terai, K., Abbas, T., Jazaeri, A. A. & Dutta, A. CRL4^{Cdt2} E3 ubiquitin ligase monoubiquitinates PCNA to promote translesion DNA synthesis. *Mol. Cell* **37**, 143–149 (2010).



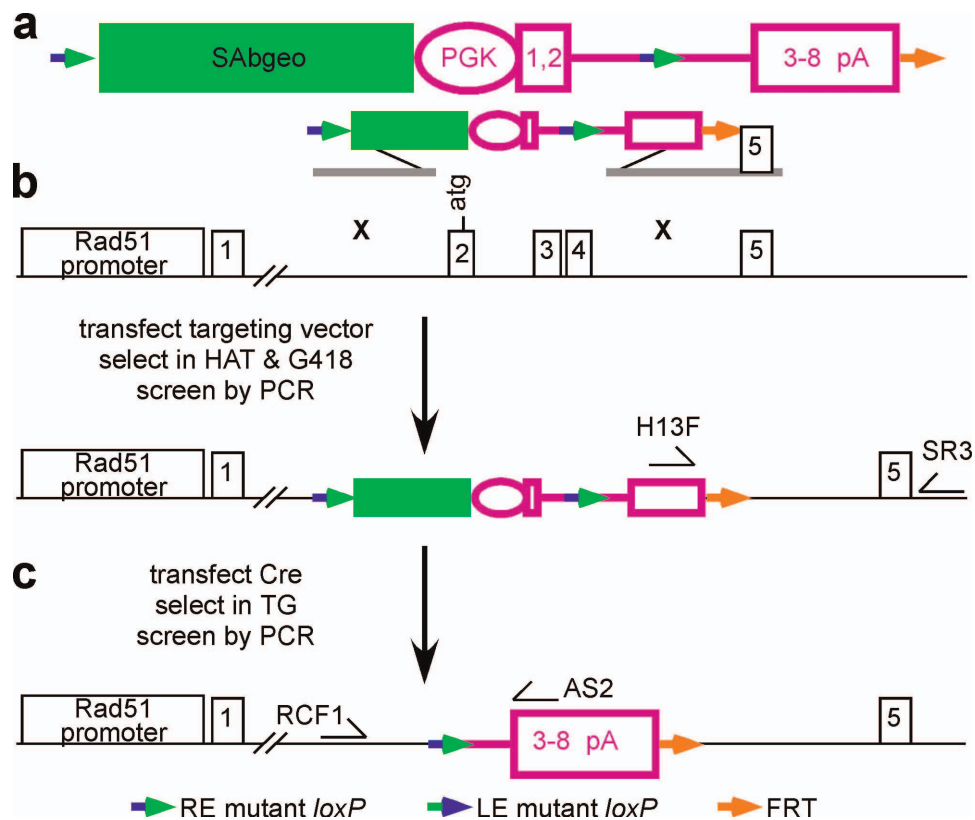
Extended Data Figure 1 | Three locations for the switch within a hairpin. There are seven mismatches located at positions 52, 111, 140, 178, 188, 204 and 246. This model shows the inverted repeats forming a hairpin to simply illustrate the location of the switch, although we do not know if hairpins form. **a**, The switch occurs at the apex of the hairpin before the first mismatch at position 52 such that the 5' MSR has the same sequence as the orange repeat.

b, The switch occurs in the stem of the hairpin after the first mismatch at position 52 but before the last mismatch at position 246 such that the 5' MSR is a mixture of both the green and orange repeat. **c**, The switch occurs at the base of the hairpin after the last mismatch in position 246 such that the 5' MSR has the same sequence as the green repeat.



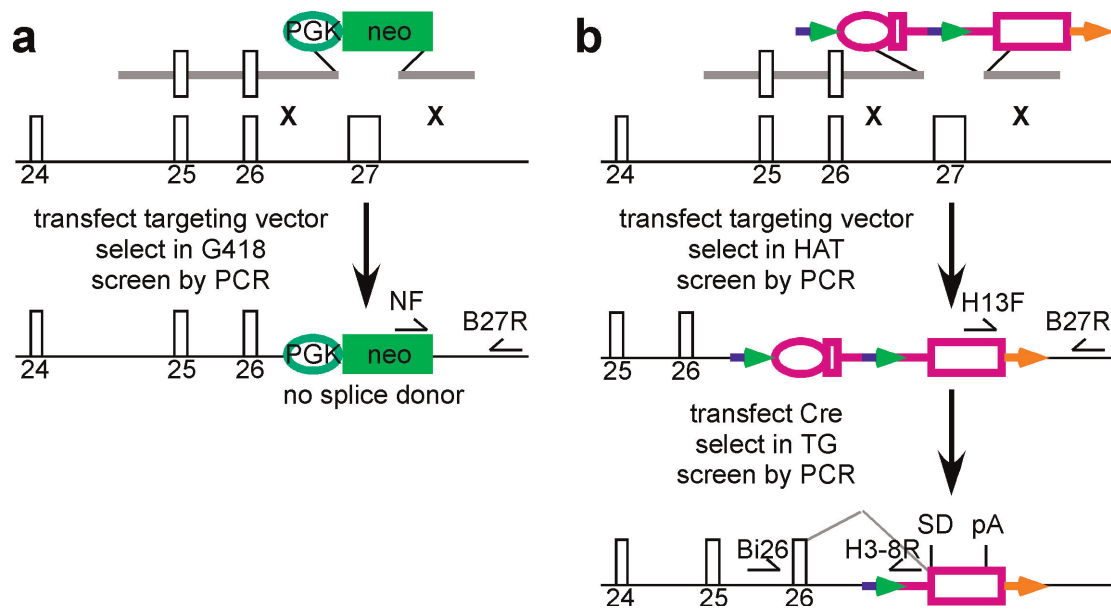
Extended Data Figure 2 | Complex chromosomal rearrangements in wild-type cells with the IRR and MRR.

a, Two-colour FISH on metaphase spreads stained with a telomeric probe (green), a MSR probe in the pericentromere (red) and counterstained with DAPI (blue). (1)–(3) Multipericentric chromosomes from cells with the IRR: (1) Typical dipericentric, (2) chromosome with extra pericentromeres and telomeres (EPT)¹⁵, (3) segmental duplication with the extra pericentromeres on only one chromatid. (4)–(8) Multipericentric chromosomes from cells with the MRR: (4) typical dipericentric, (5)–(7) EPTs, (8) extra pericentromere on only one chromatid. Chromosomal abnormalities were found for 15/19 ($P < 0.0001$, Yates-corrected chi-square test) and 18/19 ($P < 0.0001$) HAT-resistant colonies transfected with the IRR and MRR, respectively, but none were found for non-transfected cells as previously described¹⁵. **b**, Two-colour FISH on nuclei using the MMR as a probe (red) along with either chromosome 1 or 14 (green). For some nuclei the MMR associated with chromosome 14 (1) whereas for others it associated with chromosome 1 (2). Note the MMR is located to both chromosomes 14 but only one chromosome 1. Thus, the MMR moved to different altered chromosomes observed with spectral karyotyping, consistent with the notion that the MMR is the source of instability. In addition, the size of the red dot(s) varied, suggesting continuous nonallelic fusions that could expand or contract the number of MMR units. For some nuclei the MMR appeared as a discrete dot, indicating one contiguous array of reporter units (1 and 2, red insets) but for others it was speckled, suggesting arrays of MMR units were interspersed with chromosomal sequences (3, red inset). For one speckled cluster a fragment of chromosome 1 surrounded only one red dot, highlighting the complexity of this rearrangement (green inset). The MMR probe was also found protruding at the edge or outside of some nuclei, indicating these unstable structures could be extruded from the nucleus similar to micronuclei (4, red inset).



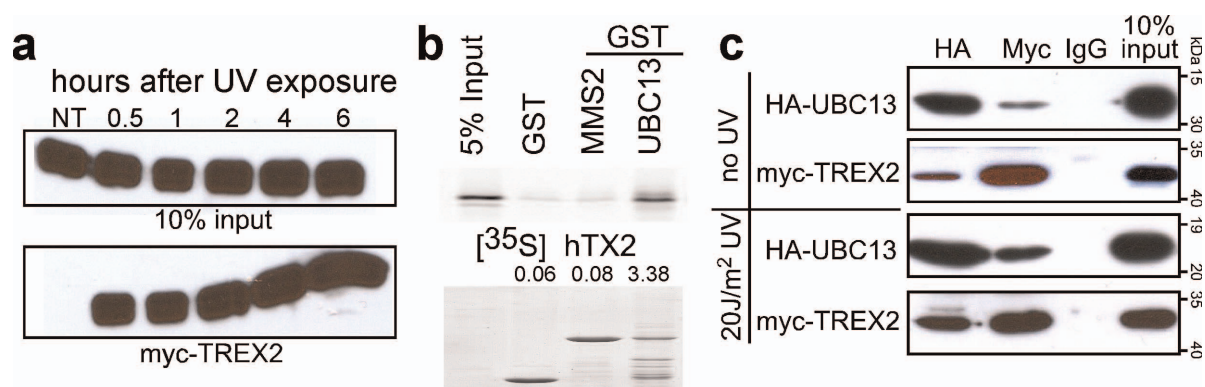
Extended Data Figure 3 | Targeting *Rad51* exons 2–4. **a**, SAβgeo-*miniHPRT* is used for selection. SAβgeo (green) is a fusion of β-galactosidase and neomycin phosphotransferase and is capable of trapping promoters to improve targeting efficiency⁴¹. A Right element (RE) mutant *loxP*¹² is in the intron (blue green arrow). In addition, another RE mutant *loxP* is 5' to SAβgeo. A FLP recombination target (FRT) is at the 3' end of *miniHPRT*^{36,43}. **b**, Replacing

Rad51 exons 2–4 (exon 2 is the first coding exon) with the SAβgeo-*miniHPRT* selection cassette. PCR is used to screen G418+HAT-resistant ES cell clones for gene targeting using primers H13F and SR3. **c**, Removal of SAβgeo, the 5' half of *miniHPRT* and a RE mutant *loxP* by Cre-mediated recombination to generate *Rad51*^{+/-Δex2-4} cells. Screen 6-thioguanine-resistant clones by PCR using primers RCF1 and AS2.



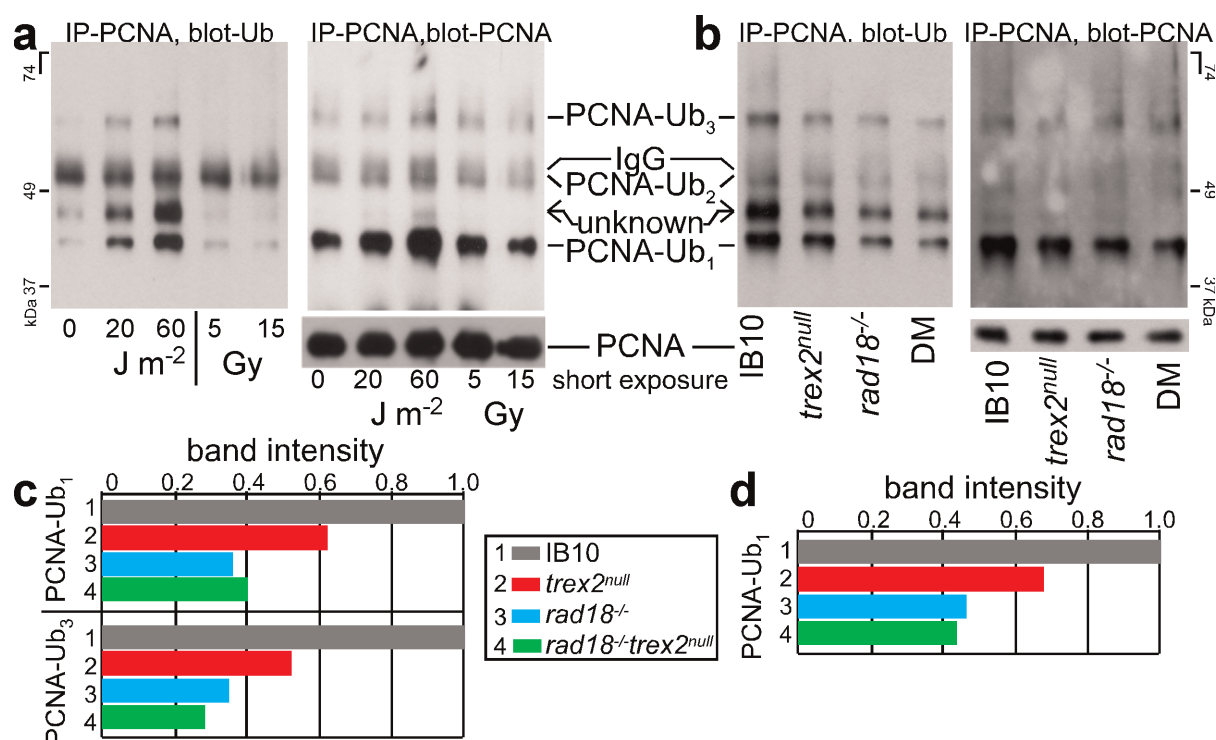
Extended Data Figure 4 | Targeting *Brca2* exon 27. There were two gene targeting vectors so we could observe cells deleted for one (*blm*^{-/-} *Brca2*^{+/Δex27-n}) and two (*blm*^{-/-} *brca2*^{Δex27-h/Δex27-n}) copies of *Brca2* exon 27. **a**, The first targeting vector (*Δex27-n*) replaced *Brca2* exon 27 with neomycin phosphotransferase (neo) and probably generated a severe defect because exon 27 was not replaced with a splice donor to ensure polyadenylation⁴⁴. This means deletion of the first copy probably caused a haploinsufficiency. The *Brca2* gene after targeting. NF and B27R are PCR primers used to screen for targeted clones. **b**, The second targeting vector (*Δex27-h*) replaced *Brca2* exon 27 with *miniHPRT* that contains a splice donor and polyadenylation sequences. Previously we showed *Brca2* exon 26 spliced into *HPRT* exon 3 to ensure polyadenylation. Cells mutated with this second

targeting vector produced a truncated BRCA2 protein at normal levels and were hypersensitive to γ -radiation and deficient in homologous recombination^{36,45,46} and replication fork maintenance⁶. Replacing the second copy of *Brca2* exon 27 with a floxed *miniHPRT*³⁶ to make *Brca2*^{Δex27-h/Δex27-n} cells. H13F and B27R primers were used to screen for targeted clones. Cre-mediated recombination removed the 5' half of *miniHPRT*. *Brca2* exon 26 splices into *miniHPRT* exons 3–8 (grey line) to generate a polyadenylated *Brca2* transcript that is deleted for exon 27^{36,45}. There is the addition of one amino acid followed by a stop codon and this transcript produces a protein at wild-type levels that associates with RAD51, presumably through the BRC motifs⁴⁶. Bi26 and H3-8R PCR primers were used to screen for Cre-mediated deletion.



Extended Data Figure 5 | TREX2's response to ultraviolet light and association with UBC13. **a**, Coimmunoprecipitation of IdU and Myc-TREX2 in HeLa cells after exposure to 20 J m⁻² ultraviolet light. No treatment, NT.

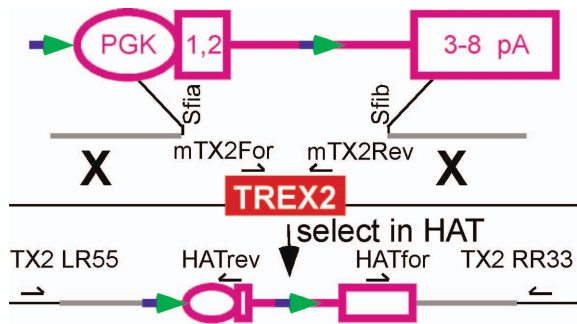
b, GST pull-down of ³⁵S-labelled short isoform wild-type (WT) TREX2²³. **c**, Coimmunoprecipitation with Myc-TREX2 and HA-UBC13 in HeLa cells before and 6 h after exposure to 20 J m⁻² ultraviolet light.



Extended Data Figure 6 | RAD18 and TREX2 ubiquitinate PCNA.

a, Exposure of AB2.2 cells to ultraviolet light, but not γ -radiation, induced PCNA ubiquitination. Immunoprecipitate endogenous PCNA and immunoblot with anti-ubiquitin (Ub, left), then strip and immunoblot with anti-PCNA (right). PCNA-Ub₁ and PCNA-Ub₃ are visible; yet, IgG obscures PCNA-Ub₂. In addition, the Ub blot, but not the PCNA blot, reveals a previously unidentified band between PCNA-Ub₁ and PCNA-Ub₂. Ultraviolet light, but not γ -radiation, increased levels of PCNA-Ub₁ and PCNA-Ub₃ as previously shown in human cells²¹ (the same was true for the unknown protein). Survival fraction: 20 J m^{-2} , 0.6%; 60 J m^{-2} , 0.06%; 5 Gy, 8%; 15 Gy; 0.001%. **b**, Analysis of *trex2^{null}* and *rad18^{-/-}* cells and double-mutant cells. In response to 60 J m^{-2} ultraviolet light, *trex2^{null}* and *rad18^{-/-}* cells had reduced levels of PCNA-Ub₁ and PCNA-Ub₃ and unknown protein as compared to IB10 cells. *rad18^{-/-}* cells exhibited a marginally greater reduction than *trex2^{null}* cells, indicating that RAD18 has a greater role in PCNA ubiquitination. The double-mutant cells failed to show a further reduction, indicating that TREX2 and RAD18 are epistatic. Some ubiquitinated PCNA was present in mutant cells, indicating that other proteins ubiquitinate PCNA; similar observations

were made for cells deleted for HLTf and SHPRH⁴⁰. For example, CRL4^{Cdt2}, independent of RAD18, monoubiquitinates PCNA with and without ultraviolet-light-induced damage⁴⁷. **c**, Bar graph illustrating the reduction of PCNA-Ub₁ and PCNA-Ub₃ in *trex2^{null}*, *rad18^{-/-}*, and double-mutant cells as shown in **b**, left (immunoprecipitation-PCNA, blot-Ub), after band intensities were quantified with ImageJ and normalized for loading with short exposure PCNA. Statistics (*t*-test) for PCNA-Ub₁ and PCNA-Ub₃ using three experiments (lanes): 1 vs 2 (0.0016, 0.0058), 1 vs 3 (0.0036, 0.0026), 1 vs 4 (0.0064, 0.0001), 2 vs 3 (0.0214, 0.0774), 2 vs 4 (0.0310, 0.0486), 3 vs 4 (0.3169, 0.1209). **d**, Bar graph illustrating the reduction of PCNA-Ub₁ in *trex2^{null}*, *rad18^{-/-}*, and double-mutant cells as shown in **b**, right (immunoprecipitation-PCNA, blot-PCNA), after band intensities were quantified with ImageJ and normalized for loading with short exposure PCNA. The stripping and re-probing leaves quantification unreliable for PCNA-Ub₃ and further work is required to clarify the extent to which Ub modification is influenced in these backgrounds. Statistics (*t*-test) for PCNA-Ub₁ using three experiments (lanes): 1 vs 2 (0.0021), 1 vs 3 (0.0061), 1 vs 4 (0.0460), 2 vs 3 (0.0212), 2 vs 4 (0.0163), 3 vs 4 (0.0604).



Extended Data Figure 7 | Deleting *Trex2* in IB10 control and *rad18*^{-/-} cells. A floxed *MiniHPRT*²⁶ was used to replace the entire *Trex2* coding sequences (located on a single exon)²⁵. Targeted clones were detected using PCR with TX2 LR55 and HATrev primers for the left arm and HATfor and TX2 RR33 primers for the right arm. Removal of the *Trex2* coding sequence was verified by PCR using mTX2For and mTX2Rev primers.

Extended Data Table 1 | IRR summary

clone	#MPS	DP	EPT	SD
1	32	3	0	0
2	32	0	0	0
3	30	3	0	0
4	33	0	0	0
5	25	1	0	0
6	31	1	0	0
7	30	0	0	0
8	23	1	0	0
9	27	2	0	0
10	24	1	0	0
11	27	0	0	0
12	31	0	2	0
13	47	0	1	0
14	53	0	2	0
15	65	2	1	0
16	46	0	1	0
17	48	1	1	0
18	49	1	0	0
19	54	0	1	1

Dipicentric (DP), extra pericentromere and telomere (EPT), metaphase spread (MPS), segmental duplication (SD).

Extended Data Table 2 | MRR summary

clone	#MPS	DP	EPT	SD
1	27	1	0	0
2	22	3	0	0
3	22	0	1	0
4	35	1	2	0
5	31	1	3	0
6	31	1	1	0
7	29	0	0	0
8	26	0	2	0
9	28	3	0	0
10	22	0	2	1
11	26	2	1	0
12	22	2	2	1
13	47	0	24	0
14	34	0	1	0
15	35	1	11	0
16	36	1	1	0
17	44	2	0	0
18	38	0	28	0
19	36	1	3	0

Dipericentric (DP), extra pericentromere and telomere (EPT), metaphase spread (MPS), segmental duplication (SD).

Extended Data Table 3 | Spectral karyotyping summary

MRR clone	nuclei	Simple EPT	Complex EPT	other
13	1	-	-	-
	2	Dup(1;1)	T(14;13)	-
	3	Dup(7;7)	T(3;13)	chromatid fusion (8;19)
	4	Dup(1;1)	T(14;13)	chromatid break (16)
	5	-	-	-
	6	-	T(14;13)	Del(1)
	7	chr1-chr4-chr17	T(14;13)	-
	8	Dup(1;1) Dup(7;7)	T(14;13) T(2;?)	Del(3)
	9	Dup(1;1)	-	-
15	1	Dup(1;1)	-	T(11;14) T(14;11)
	2	Dup(1;1)(1;)	-	T(11;14) T(14;13) Del(13)
	3	Dup(1;1)	T(?;13)	T(11;14) T(14;11)
	4	Dup(1;1)	-	T(11;3)
	5	Dup(1;1)	-	Ins(3;11) T(11;3) Del(3)
	6	Dup(1;1)	-	T(11;14) T(14;11)
	7	Dup(1;1)	-	T(11;14) T(14;11)
18	1	Dup(1;1) Dup(7;7)	T(14;11)	-
	2	-	T(14;11)	-
	3	Dup(1;1) Dup(12;12)	T(14;11)	-
	4	Dup(1;1)	T(14;11)	-
	5	-	T(14;11)	-
	6	Dup(1;1) Dup(7;7)	T(14;11)	-
	7	Dup(1;1) Dup(7;7)	T(14;11)	chromatid fusion (17;14)
	8	Dup(1;1) Dup(12;12)	T(14;11)	-

Simple extra pericentromere and telomere (EPT) involves one chromosome. Complex EPT involves more than one chromosome. Other has only one pericentromere.

A two-domain elevator mechanism for sodium/proton antiport

Chiara Lee¹, Hae Joo Kang¹, Christoph von Ballmoos², Simon Newstead^{1†}, Povilas Uzdavinyas², David L. Dotson³, So Iwata^{1,4,5}, Oliver Beckstein³, Alexander D. Cameron^{1,4,5,6} & David Drew^{1,2}

Sodium/proton (Na^+/H^+) antiporters, located at the plasma membrane in every cell, are vital for cell homeostasis¹. In humans, their dysfunction has been linked to diseases, such as hypertension, heart failure and epilepsy, and they are well-established drug targets². The best understood model system for Na^+/H^+ antiport is NhaA from *Escherichia coli*^{1,3}, for which both electron microscopy and crystal structures are available^{4–6}. NhaA is made up of two distinct domains: a core domain and a dimerization domain. In the NhaA crystal structure a cavity is located between the two domains, providing access to the ion-binding site from the inward-facing surface of the protein^{1,4}. Like many Na^+/H^+ antiporters, the activity of NhaA is regulated by pH, only becoming active above pH 6.5, at which point a conformational change is thought to occur⁷. The only reported NhaA crystal structure so far is of the low pH inactivated form⁴. Here we describe the active-state structure of a Na^+/H^+ antiporter, NapA from *Thermus thermophilus*, at 3 Å resolution, solved from crystals grown at pH 7.8. In the NapA structure, the core and dimerization domains are in different positions to those seen in NhaA, and a negatively charged cavity has now opened to the outside. The extracellular cavity allows access to a strictly conserved aspartate residue thought to coordinate ion binding^{1,8,9} directly, a role supported here by molecular dynamics simulations. To alternate access to this ion-binding site, however, requires a surprisingly large rotation of the core domain, some 20° against the dimerization interface. We conclude that despite their fast transport rates of up to 1,500 ions per second³, Na^+/H^+ antiporters operate by a two-domain rocking bundle model, revealing themes relevant to secondary-active transporters in general.

Na^+/H^+ antiporters are secondary active transporters that are conserved across all biological kingdoms to maintain the internal pH, cell volume and sodium concentration of the cell¹, a mechanism first proposed by West and Mitchell¹⁰. Na^+/H^+ antiporters are members of the large monovalent cation proton antiporter (CPA) superfamily that includes, among others, the CPA1 and CPA2 clades². It is generally thought that Na^+/H^+ antiporters from the CPA1 clade catalyse electro-neutral sodium-proton exchange (SLC9A1–9 in mammals (also known as NHE1–9)), whereas CPA2 members are thought to be electrogenic (SLC9B1–2 in mammals (also known as NHA1–2))², with stoichiometries of $2\text{H}^+ : 1\text{Na}^+$ and $3\text{H}^+ : 2\text{Na}^+$ ions reported^{11,12}. Like all secondary-active transporters, Na^+/H^+ antiporters are thought to operate by an alternating access mechanism; however, the different conformational states of the transport cycle have yet to be determined. The inward-facing structure of the well-characterized bacterial CPA2 protein *E. coli* NhaA is the only representative crystal structure⁴.

Using fluorescent-based methods^{13,14}, we screened members of the CPA2 clade² for their suitability for structural studies. NapA from *T. thermophilus*, which has 21% sequence identity to human NHA2, was thus identified (see Methods and Supplementary Figs 1a and 2). Although the overall sequence homology to the better characterized

E. coli NhaA is low, <15% identity, residues identified to be important for transport in NhaA and mammalian homologues are nonetheless well conserved (Supplementary Fig. 2). Using isolated inside-out membrane vesicles it was previously shown that *T. thermophilus* NapA is active above pH 6, with maximum activity for sodium at pH 8 (ref. 15), similar to *E. coli* NhaA¹⁶. To measure activity in an isolated system, we co-reconstituted purified NapA and *E. coli* F_1F_0 ATP synthase into liposomes (see Methods). After establishment of a pH gradient by the addition of ATP, proton efflux was monitored in response to Na^+ or Li^+ addition (Fig. 1a, b). In this experimental set-up, the apparent Michaelis constant (K_m) values for Na^+ or Li^+ were determined to be 4.0 ± 0.3 (mean \pm s.d.) or 0.41 ± 0.04 mM, respectively, similar to the affinities from inside-out vesicles¹⁵ (Fig. 1c). In experiments with NapA proteoliposomes trapped with a water-soluble pH sensitive dye, dissipation of the membrane potential ($\Delta\psi$) stimulated exchange activity in the presence of a Na^+ gradient, confirming the electrogenic nature of NapA (Fig. 1d, bottom). Furthermore, in the absence of a H^+ or Na^+ gradient, NapA transport activity was solely driven by $\Delta\psi$ (Fig. 1d, top). Taken together, *T. thermophilus* NapA has a similar antiport profile to *E. coli* NhaA, and consequently functional and structural studies of the two proteins can complement one another.

The structure of NapA at pH 7.8 was solved by multiple isomorphous replacement with anomalous scattering in combination with multi-crystal averaging. The highest resolution data correspond to a triple mutant of NapA, in which three cysteine residues, which have no apparent effect on functional activity, were introduced to facilitate phasing (see Methods, Supplementary Tables 1 and 2 and Supplementary Fig. 1b). The structure was refined at a resolution of 3 Å, R_{factor} value of 22.3% and R_{free} value of 24.8% (Supplementary Table 1, Supplementary Fig. 3 and Methods). Electron density maps at 3.7 Å of wild-type NapA at pH 9.0 show no clear structural differences between the wild-type and the triple cysteine mutant.

NapA is built from 13 transmembrane (TM) helices with an $\text{N}_{\text{out}}-\text{C}_{\text{in}}$ topology (Fig. 2 and Supplementary Fig. 4). Relative to NhaA, it has an additional helix at the amino terminus. To facilitate comparison to NhaA, we refer to the first helix as TM–1 (Fig. 2a and Supplementary Fig. 4). TM–1 to TM5 and TM7 to TM12 are topologically similar but oppositely orientated in the plane of the membrane (Supplementary Fig. 4). These six-transmembrane-helice topology inverted repeats intertwine to form a core (ion-translocation) and dimerization (interface) domain, and are linked together by TM6 (Fig. 2b and Supplementary Fig. 4). The NapA and NhaA⁴ structures are very similar (Fig. 2b). As there is a change in the position of the core relative to the dimerization domain, however, the similarity is best seen when the two domains are superposed separately; root mean squared deviation (r.m.s.d.) of 1.8 Å for 134 out of 148 pairs of C α atoms of the core domain transmembrane helices, and 1.9 Å for 62 out of 88 of the dimerization domain transmembranes (Supplementary Fig. 5a, b). The change in position of

¹Division of Molecular Biosciences, Imperial College London, London SW7 2AZ, UK. ²Centre for Biomembrane Research, Department of Biochemistry and Biophysics, Stockholm University, SE-106 91 Stockholm, Sweden. ³Center for Biological Physics, Department of Physics, Arizona State University, Tempe, Arizona 85287-1504, USA. ⁴Membrane Protein Laboratory, Diamond Light Source, Harwell Science and Innovation Campus, Didcot, Chilton, Oxfordshire OX11 0DE, UK. ⁵Research Complex at Harwell Rutherford, Appleton Laboratory, Harwell, Oxford, Didcot, Oxfordshire OX11 0FA, UK. ⁶School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK. [†]Present address: Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK.

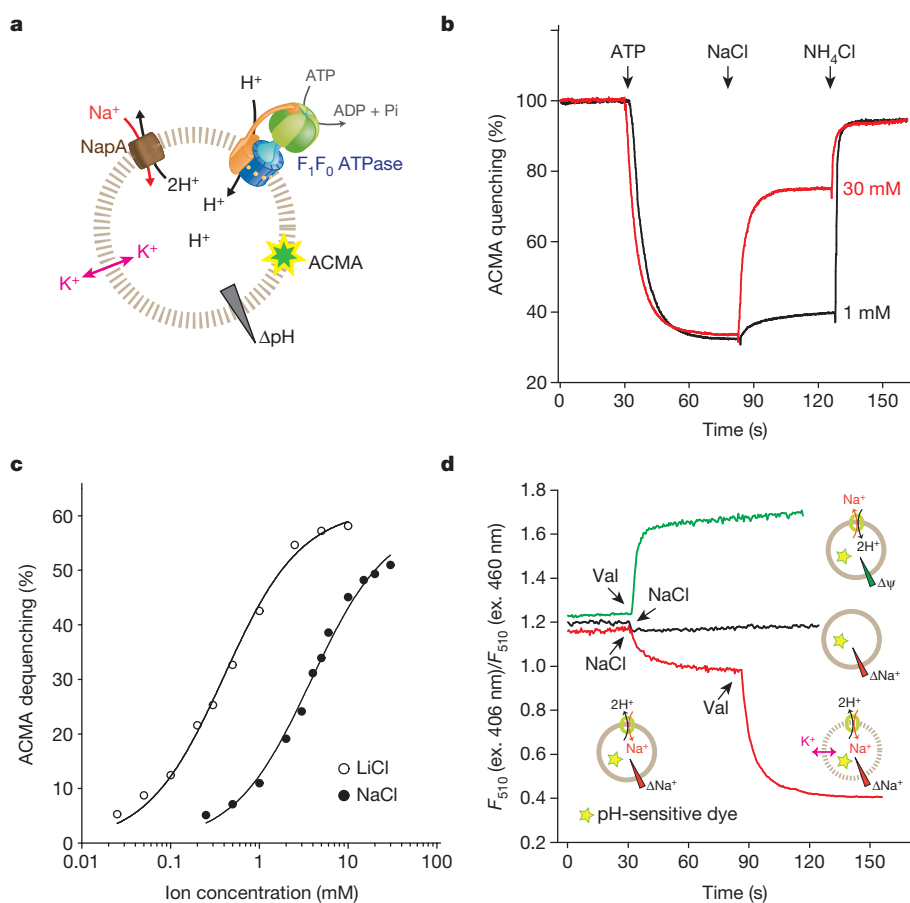


Figure 1 | NapA Na⁺/H⁺ transport activity is electrogenic. **a**, Experimental set-up for determination of Na⁺/Li⁺ affinity and electrogenicity (the most likely stoichiometry is shown). The ATP synthase and NapA are co-reconstituted in liposomes. Free K⁺ diffusion by valinomycin suppresses the effect of Δψ. ACMA, 9-amino-6-chloro-2-methoxyacridine. **b**, Representative ACMA fluorescence traces for Na⁺/H⁺ antiporter activity. ATP-driven H⁺ pumping establishes a ΔpH (acidic inside) as monitored by quenching of fluorescence. H⁺ efflux is initiated by the addition of NaCl/LiCl, and further NH₄Cl addition collapses the proton gradient. **c**, Apparent binding affinity for Na⁺ (filled circles) and Li⁺ (open circles) in NapA (pH 7.5). **d**, Na⁺/H⁺ antiporter activity monitored with liposome entrapped pH sensitive fluorophore pyranine. The fluorescence was recorded at 510 nm (excitation 406 and 460 nm) and the ratio of the two values was plotted against time. Na⁺-gradient (~120 mV)-driven Na⁺/H⁺ exchange was initiated by the addition of NaCl (red). Addition of valinomycin at 90 s leads to further H⁺ efflux; this releases the inhibitory membrane potential established during the first transport phase. No Na⁺-gradient-driven transport was observed in liposomes without NapA (black). Δψ-driven Na⁺/H⁺-exchange transport was initiated by establishing a K⁺/valinomycin diffusion potential (green). No activity was observed in liposomes without NapA (data not shown). ex., excitation; F, fluorescence.

the core relative to the dimerization domain, which may reflect the difference in the pH at which the transporters were crystallized, gives rise to a large negatively charged cavity that is now open to the outside in contrast to the inward-facing funnel seen in NhaA (Figs 2b and 3b). The interaction between TM2 and TM4 and TM5 tightly closes the cytoplasmic side of the cavity.

NapA, like other bacterial and mammalian Na⁺/H⁺ antiporters^{6,17,18}, purifies as a dimer and is clearly dimeric in the structure with an extensive interface burying a surface area of 1,800 Å² (Fig. 3 and Supplementary Fig. 1c). In molecular dynamics simulations, the dimer sits entirely within a model membrane bilayer (Fig. 3c). The backbone r.m.s.d. from the starting conformation of each monomer increased to about 2.0 Å over the time course of the simulation (Supplementary Fig. 6), indicating a slow relaxation of the crystal structure in the native-like membrane environment. The dimer has a crystallographic two-fold operator approximately parallel to the transmembrane helices, and it is largely made up of tight hydrophobic helix-helix packing between TM-1 on one monomer and TM7 on the other. There are also contacts between the ends of TM2 and TM9 (Fig. 3a, b). NapA lacks the β-hairpin domain that makes most of the contact between protomers at the extracellular membrane surface in NhaA^{4,19}. The dimer interface in NapA more closely resembles the dimer interface modelled in the 7 Å electron crystallography structure of NhaP1 (ref. 20), a CPA1 member from *Methanocaldococcus jannaschii* (formerly *Methanococcus jannaschii*). NhaP1 also has 13 transmembrane helices built up from two six-transmembrane topology repeat units.

The substrate-binding cavity is open to the extracellular side. The cavity begins above the dimerization interface and funnels between the dimerization and core domains, ending in the middle of the membrane (Fig. 3). It is considerably more open than the inward-facing cavity of NhaA⁴ and is negatively charged, being lined with glutamate residues. Although there is a similar distribution of charged residues in NhaA

and NapA, few, including those that have been predicted to be involved in pH sensing¹, are conserved between the two proteins (Supplementary Fig. 7). Near the base of the cavity are two highly conserved aspartates, Asp 156 and Asp 157, located on TM5 (Supplementary Figs 2 and 7). The residues equivalent to these aspartates in NhaA, Asp 163 and Asp 164 (Supplementary Figs 2 and 7), probably coordinate ion binding based on their position, conservation with mammalian Na⁺/H⁺ antiporters², phenotypes of mutants^{1,21}, isothermal titration calorimetry (ITC) experiments⁹ and molecular dynamics simulations⁸. In NapA, mutation of either residue to alanine¹⁵ or asparagine abolishes transport activity completely (Supplementary Fig. 1d and Supplementary Table 2). With NapA crystals grown at pH 7.8, the aspartate residues are likely to be deprotonated. Consistent with this, Asp 157 is orientated towards the centre of the cavity rather than hydrogen bonding with the backbone of TM4, as seen for Asp 164 in NhaA, an interaction that requires Asp 164 to be protonated (Supplementary Fig. 5c). Using solid-state membrane electrophysiology it was previously shown that NhaA transports cations equally well in either direction and the transport activity profile fits a simple H⁺ versus Na⁺ kinetic binding model to a single common site²². To investigate whether Na⁺ ions would bind as expected, we carried out equilibrium molecular dynamics simulations of outward-facing NapA in a model membrane bilayer. Simulations were carried out with both Asp 157 and Asp 156 deprotonated, as they are likely to be in the crystal structure, and also as combinations of their neutral and charged forms. With both aspartates charged, Na⁺ ions spontaneously entered the negatively charged extracellular cavity to bind to Asp 157 (Fig. 3c and Supplementary Fig. 8). Na⁺ ions were concentrated at Asp 157 (Supplementary Fig. 8) and multiple distinct binding and unbinding events could be observed, which is qualitatively consistent with weak binding. By contrast, Na⁺ ion binding was not observed when Asp 157 was protonated and was markedly reduced when Asp 156 was neutral (Supplementary Fig. 8). In molecular dynamics simulations of

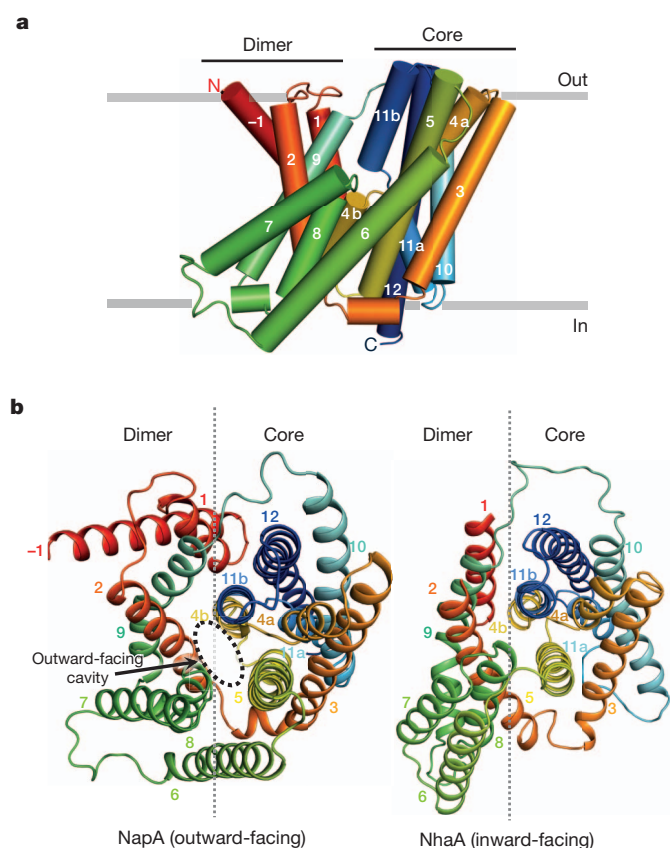


Figure 2 | Outward-facing NapA structure. **a**, Cartoon representation of NapA as viewed in the plane of the membrane. Transmembrane helices –1 to 12 have been coloured from red at the amino terminus to blue at the carboxy terminus, with the position of the membrane depicted in grey. **b**, Outward-facing NapA structure (left) and inward-facing NhaA structure (right) as viewed from the extracellular side with the removal of the β -hairpin domain located between transmembrane helices 1 and 2 from NhaA to facilitate visual comparison. The dimerization and core domain boundaries are represented by a grey line.

inward-facing NhaA, Na^+ binds to the equivalent aspartate (Asp 164)⁸, which is positioned at the base of the cytoplasmic cavity.

Asp 156 and Asp 157 on TM5 are located next to the point at which the two antiparallel discontinuous helices, TM4a–b and TM11a–b, cross over in the core domain (Supplementary Fig. 5c). Although discontinuous helices are a common feature for ion-binding in transporters²³, the antiparallel crossing over of transmembrane helices is a unique feature of the NhaA fold^{4,24}. It results in the positive dipole ends of TM4a and TM11a facing one another and likewise the negative dipole ends of TM4b and TM11b. In NhaA, the dipoles are proposed to be neutralized by the side chains of an aspartate (Asp 133; TM4a–b) and a lysine (Lys 300; TM10), respectively (Supplementary Fig. 5c). The lysine (Lys 305) is retained in NapA but the aspartate is replaced by a serine, as it is in human NHA2 (Supplementary Figs 5c and 7). From the structure, however, it appears the side chain of Glu 333 from TM11b can take a similar position to the carboxylate of Asp 133 (Supplementary Figs 5c and 7). The positions of these two residues are pseudo-symmetrically related; the functional significance of this swap is unclear. Although the mutation of Glu 333 to alanine affected the apparent affinity for Li^+ only slightly (<3-fold), the affinity for Na^+ was severely decreased (>15-fold), which is in agreement with the results obtained from an Asp 133 to alanine mutation in the NhaA protein²⁵ (Supplementary Table 2). By contrast, the mutation of the highly conserved Lys 305 to alanine severely decreased both Na^+ and Li^+ affinity (>20-fold; Supplementary Table 2). It was recently speculated that Lys 300 in NhaA could have more than a stabilizing role and that it could also form part of the pH activation mechanism, as mutation to arginine

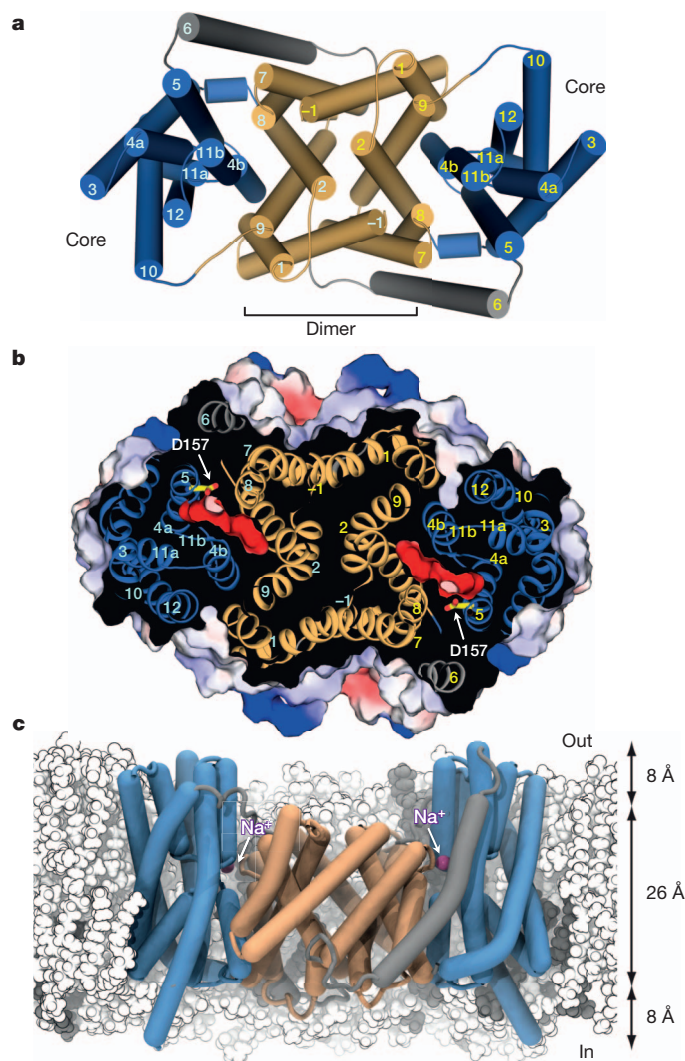


Figure 3 | Structure of NapA dimer and proposed Na^+/Li^+ -binding site. **a**, Cartoon representation of the NapA dimer structure, with transmembrane helices for one monomer labelled in cyan and the other in yellow. The transmembrane helices of the dimerization domains are coloured in pale orange, and the core ion-translocation domains are in sky blue, connected together by TM6 shown in grey. **b**, Electrostatic surface representation showing the location of the negatively charged extracellular cavity as a section through the protein from above (same orientation as in **a**). The proposed ion-binding aspartate, Asp 157, is illustrated as a stick model. **c**, In molecular dynamics simulations, the NapA dimer is stable in a 4 POPE (white):1 POPG (grey) lipid membrane (POPE, 1-palmitoyl-2-oleoylphosphatidylethanolamine; POPG, 1-palmitoyl-2-oleoylphosphatidylglycerol). Sodium ions (magenta spheres) spontaneously bind to Asp 157 at the bottom of the outward-facing cavity. The approximate thickness of the hydrophobic core and headgroup regions of the membrane are indicated.

changed the pH at which NhaA becomes active⁹. In the NapA structure, Lys 305 forms a salt bridge with Asp 156 (Supplementary Figs 5c and 8a). Notably, we have recently observed a similar interaction in a different crystal form of NhaA at low pH (O.B. *et al.*, manuscript in preparation). We would expect, however, that after cation binding to Asp 156 and Asp 157 the salt bridge is disrupted. As such, the interaction observed here supports the role of lysine in pH activation.

Previously, it was proposed that the exceptionally fast transport of Na^+/H^+ antiporters would primarily involve local rearrangements of the finely electrostatically balanced discontinuous helices in the core domain¹⁴. However, the major structural difference between NapA and NhaA is in the position of the core domain in relation to the dimerization domain (Supplementary Fig. 5a, b). With reference to the dimerization

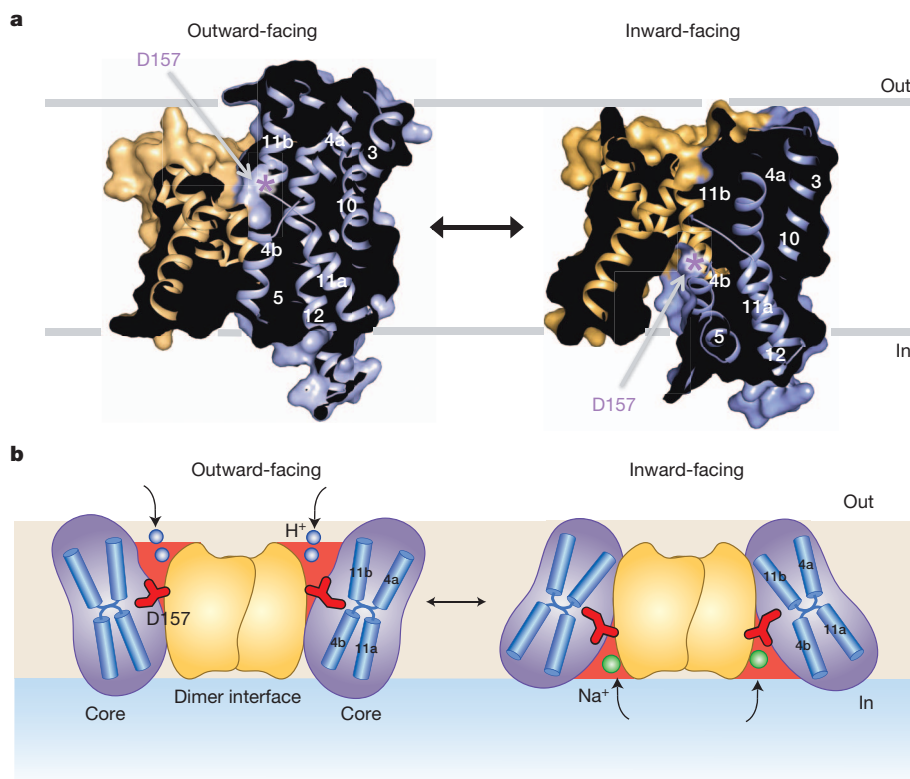


Figure 4 | Alternating access model of sodium-proton antiporter. **a**, Surface representation showing a section through the outward-facing NapA structure (left) and inward-facing NapA model (right) (see Methods and Supplementary Videos 1 and 2). The position of Asp 157 is denoted by an asterisk, and the helices have been coloured as in Fig. 3. For the sake of clarity, only one molecule is shown. **b**, Schematic of the proposed transport mechanism that illustrates the conformational changes with the core moving against the dimerization domain. Protons (shown in blue) bind to the core domain in the outward-facing state (left) causing it to switch to the inward-facing state (right). On the inside, protons are exchanged for sodium (green) and the core domain moves back to the outside. Asp 157 (shown in red), which is crucial for binding both one of the protons and the sodium ion, moves approximately 10 Å during this process. Other residues involved in ion binding are not shown.

interface, which seems likely to remain stable during transport across the membrane^{19,26}, the core domain in the NhaA structure is rotated by 21° relative to the core of NapA (Fig. 4a, b). This large rotation of the core domain closes the cavity seen on the outside of NapA and opens the cytoplasmic funnel on the inside, as observed in NhaA (Supplementary Videos 1 and 2). During this process, the two cation binding aspartates, which are in line with the tip of TM8 in the dimerization domain of the outward-facing NapA structure, are shifted 10 Å towards the cytoplasmic surface of the transporter (Fig. 4a). This elevator movement of a substrate-binding domain, in this case to carry Na⁺ (Li⁺) ions from one side of the membrane to the other in exchange for protons, resembles that of the transport mechanism seen in the glutamate transporter Glt_{Ph} (ref. 27). A two-domain transport mechanism was also recently predicted in NhaA, on the basis of a consideration of the two symmetry-related inverted repeats as well as elastic network models and biochemical cross-linking²⁸. Such a mechanism was also proposed for the bile acid sodium symporter ASBT_{NM} (ref. 24), which is a structural homologue of NhaA and NapA.

In summary, the structure of NapA is consistent with a single ion-translocation site mechanism, of which the strictly conserved Asp 157 (Asp 164 in NhaA) is ideally positioned for binding ions in outward- and inward-facing states (Fig. 4b). To provide alternating access to the aspartate, however, requires a surprisingly large movement of the core domain that twists around the dimerization interface, in essence creating a two-domain rocking bundle model. Although further structures are needed to clarify how ion binding and release is coupled to these global changes, the NapA structure provides a clear example that the extent of the conformational change may not necessarily correlate with the rate of transport or the size of the substrate transported, as previously assumed. Thus, the NapA structure also reveals fresh mechanistic insights relevant to all types of ion-coupled transporters.

METHODS SUMMARY

NapA was cloned into a cleavable green fluorescent protein (GFP)–His₈ fusion vector pWaldoGFP¹³. The fusion protein was expressed in *E. coli*, solubilized in 1% dodecyl-β-D-maltopyranoside (DDM), and purified to homogeneity in either

DDM or 1% nonyl-β-D-maltopyranoside. The NapA proteoliposome uptake assay was modified as previously described³. Crystals were grown at either pH 7.8 or 9.0 by the vapour diffusion method. Data were collected on beamlines I02 and I03 at the Diamond Light Source or ID 23-1 and 23-2 beamlines at the European Synchrotron Radiation Facility. The protein was derivatized before crystallization by incubation with 2.5 mM mercury acetate. The NapA structure was solved by multiple isomorphous replacement with anomalous scattering in combination with multi-crystal averaging and refined at a resolution of 3 Å. Molecular dynamics simulations were performed as described in the Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 4 February; accepted 17 July 2013.

Published online 1 September 2013.

- Padan, E. The enlightening encounter between structure and function in the NhaA Na⁺-H⁺ antiporter. *Trends Biochem. Sci.* **33**, 435–443 (2008).
- Brett, C. L., Donowitz, M. & Rao, R. Evolutionary origins of eukaryotic sodium/proton exchangers. *Am. J. Physiol. Cell Physiol.* **288**, C223–C239 (2005).
- Taglicht, D., Padan, E. & Schuldiner, S. Overproduction and purification of a functional Na⁺/H⁺ antiporter coded by *nhaA* (*ant*) from *Escherichia coli*. *J. Biol. Chem.* **266**, 11289–11294 (1991).
- Hunte, C. *et al.* Structure of a Na⁺/H⁺ antiporter and insights into mechanism of action and regulation by pH. *Nature* **435**, 1197–1202 (2005).
- Williams, K. A., Geldmacher-Kaufer, U., Padan, E., Schuldiner, S. & Kuhlbrandt, W. Projection structure of NhaA, a secondary transporter from *Escherichia coli*, at 4.0 Å resolution. *EMBO J.* **18**, 3558–3563 (1999).
- Williams, K. A. Three-dimensional structure of the ion-coupled transport protein NhaA. *Nature* **403**, 112–115 (2000).
- Kozachkov, L. & Padan, E. Conformational changes in NhaA Na⁺/H⁺ antiporter. *Mol. Membr. Biol.* **30**, 90–100 (2012).
- Arkin, I. *et al.* Mechanism of Na⁺/H⁺ antiporting. *Science* **317**, 799–803 (2007).
- Maes, M., Rimon, A., Kozachkov-Magrisso, L., Friedler, A. & Padan, E. Revealing the ligand binding site of NhaA Na⁺/H⁺ antiporter and its pH dependence. *J. Biol. Chem.* **287**, 38150–38157 (2012).
- West, I. C. & Mitchell, P. Proton/sodium ion antiport in *Escherichia coli*. *Biochem. J.* **144**, 87–90 (1974).
- Taglicht, D., Padan, E. & Schuldiner, S. Proton-sodium stoichiometry of NhaA, an electrogenic antiporter from *Escherichia coli*. *J. Biol. Chem.* **268**, 5382–5387 (1993).
- Pinner, E., Padan, E. & Schuldiner, S. Kinetic properties of NhaB, a Na⁺/H⁺ antiporter from *Escherichia coli*. *J. Biol. Chem.* **269**, 26274–26279 (1994).
- Drew, D., Lerch, M., Kunji, E., Slotboom, D. J. & de Gier, J. W. Optimization of membrane protein overexpression and purification using GFP fusions. *Nature Methods* **3**, 303–313 (2006).

14. Sonoda, Y. *et al.* Benchmarking membrane protein detergent stability for improving throughput of high-resolution X-ray structures. *Structure* **19**, 17–25 (2011).
15. Furrer, E. M., Ronchetti, M. F., Verrey, F. & Pos, K. M. Functional characterization of a NapA Na⁺/H⁺ antiporter from *Thermus thermophilus*. *FEBS Lett.* **581**, 572–578 (2007).
16. Padan, E. *et al.* NhaA of *Escherichia coli*, as a model of a pH-regulated Na⁺/H⁺ antiporter. *Biochim. Biophys. Acta* **1658**, 2–13 (2004).
17. Rimón, A., Tzuber, T. & Padan, E. Monomers of the NhaA Na⁺/H⁺ antiporter of *Escherichia coli* are fully functional yet dimers are beneficial under extreme stress conditions at alkaline pH in the presence of Na⁺ or Li⁺. *J. Biol. Chem.* **282**, 26810–26821 (2007).
18. Hisamitsu, T., Ben Ammar, Y., Nakamura, T. Y. & Wakabayashi, S. Dimerization is crucial for the function of the Na⁺/H⁺ exchanger NHE1. *Biochemistry* **45**, 13346–13355 (2006).
19. Appel, M., Hizlan, D., Vinothkumar, K. R., Ziegler, C. & Kuhlbrandt, W. Conformations of NhaA, the Na⁺/H⁺ exchanger from *Escherichia coli*, in the pH-activated and ion-translocating states. *J. Mol. Biol.* **388**, 659–672 (2009).
20. Goswami, P. *et al.* Structure of the archaeal Na⁺/H⁺ antiporter NhaP1 and functional role of transmembrane helix 1. *EMBO J.* **30**, 439–449 (2011).
21. Kuwabara, N., Inoue, H., Tsuboi, Y., Nakamura, N. & Kanazawa, H. The fourth transmembrane domain of the *Helicobacter pylori* Na⁺/H⁺ antiporter NhaA faces a water-filled channel required for ion transport. *J. Biol. Chem.* **279**, 40567–40575 (2004).
22. Mager, T., Rimón, A., Padan, E. & Fendler, K. Transport Mechanism and pH Regulation of the Na⁺/H⁺ Antiporter NhaA from *Escherichia coli*. *J. Biol. Chem.* **286**, 23570–23581 (2011).
23. Screpanti, E. & Hunte, C. Discontinuous membrane helices in transport proteins and their correlation with function. *J. Struct. Biol.* **159**, 261–267 (2007).
24. Hu, N. J., Iwata, S., Cameron, A. D. & Drew, D. Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT. *Nature* **478**, 408–411 (2011).
25. Noumi, T., Inoue, H., Sakurai, T., Tsuchiya, T. & Kanazawa, H. Identification and characterization of functional residues in a Na⁺/H⁺ antiporter (NhaA) from *Escherichia coli* by random mutagenesis. *J. Biochem.* **121**, 661–670 (1997).
26. Vinothkumar, K. R., Smits, S. H. & Kuhlbrandt, W. pH-induced structural change in a sodium/proton antiporter from *Methanococcus jannaschii*. *EMBO J.* **24**, 2720–2729 (2005).
27. Reyes, N., Ginter, C. & Boudker, O. Transport mechanism of a bacterial homologue of glutamate transporters. *Nature* **462**, 880–885 (2009).
28. Schushan, M. *et al.* A model-structure of a periplasm-facing state of the NhaA antiporter suggests the molecular underpinnings of pH-induced conformational changes. *J. Biol. Chem.* **287**, 18249–18261 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to D. Slotboom for critical reading of the manuscript and N.-J. Hu for assistance in data collection. Data were collected at the European Synchrotron Radiation Facility and Diamond Light Source, with excellent assistance from beamline scientists. This work was funded by grants from the Medical Research Council (MRC grant G0900990 to A.D.C. and D.D.), the Swedish Research Council (to C.v.B. and D.D.) and the BBSRC (BB/G02325/1 to S.I.). The authors are grateful for the use of the Membrane Protein Laboratory funded by the Wellcome Trust (grant 062164/Z/00/Z) at the Diamond Light Source Limited and The Centre for Biomembrane Research (CBR), supported by the Swedish Foundation for Strategic Research. Computer simulations were partially run on XSEDE resources (grant TG-MCB120151 to O.B.). C.L. was a recipient of a BBSRC-funded PhD scholarship, H.J.K. a Human Frontiers Science Program (HFSP) postdoctoral fellowship, and D.D. acknowledges the support from The Royal Society through the University Research Fellow (URF) scheme.

Author Contributions A.D.C. and D.D. designed the project. Cloning, expression screening, protein purification and crystallization were carried out by C.L. and D.D. with assistance from H.J.K., S.N., S.I. and A.D.C. Data collection and structural determination were carried out by C.L., D.D. and A.D.C. Experiments for functional analysis were designed by C.v.B. and D.D. and carried out by C.v.B., C.L., P.U. and D.D. Molecular dynamics simulations were carried out by D.L.D. and O.B. A.D.C. and D.D. wrote the manuscript with contributions from C.L., H.J.K., C.v.B. and O.B.

Author Information The coordinates and the structure factors for NapA have been deposited in the Protein Data Bank under accession code 4BWZ. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.D.C. (a.cameron@warwick.ac.uk) or D.D. (d.drew@imperial.ac.uk; ddrew@dbb.su.se).

METHODS

Thermus thermophilus NapA sequence (Uniprot accession number Q72IM4); residues progressively substituted to cysteine are underlined, and additional C-terminal residues retained after TEV cleavage are shown in italics (see next section for cloning details).

MHGAHLLLEIFYLLAAQVMAFIFKRLNQPVVIGEVLAGVLGVPALLGLV
HEGEILEFLAELGAVFLFMVGLLETRLKDILAVGKEAFLVAVLGVALPFLGGY
LYGLEIGFETLPALFLGTALVATSVGITARVLQELGVLSRPYSRIILGAAVIDDV
LGLIVLAVVNGVAETGQVEVGAIITRLIVLSVVFGVLAFLSTLIARLPLRLPV
GSPGLFAALGVGMAALAASIGLAPIVGAFLGGMILSEVREKYRLPEPIFAIES
FLAIPFFAMVGVRLLELSALASPVVLVAGTVTVIAILGKVLGGFLGALTQGV
RSALTGVGVGMAPRGEVGLIVAALGLKAGAVNEEYAIVLFMVVFTTLFAFP
ALKPLIAWTERERAAKEGSENLVYQ.

Expression screening, mutagenesis, protein purification and characterization.

NhaA homologues were cloned as green fluorescent protein (GFP)–His₈ fusions into the vector pWaldGFPe (ref. 13), as fluorescence from the C-terminal GFP fusion is a reliable reporter of membrane-integrated expression²⁹. The monodispersity of expressed fusions were screened in a number of different detergents by fluorescence-detection size-exclusion chromatography³⁰. NapA from *T. thermophilus* was selected as a suitable candidate showing stability in a wide range of detergents including the harsh detergent *N*-dodecyl-*N,N*-dimethylamine-*N*-oxide¹⁴. Expression levels of the protein were initially low in standard culture conditions, but improved significantly using MemStar, which is a new strategy for boosting expression of membrane proteins in *E. coli* (C.L. *et al.*, manuscript in preparation). In brief, Lemo21(DE3) cells³¹ were grown at 37 °C in PASM-5052 media³², with and without selenomethionine incorporation, and induced with 0.4 mM isopropyl-β-D-thiogalactoside (IPTG) at an absorbance (*A*_{600 nm}) of 0.5 for overnight incubation at 25 °C.

Wild-type NapA and mutants generated by Quickchange protocol (Agilent Technologies) were purified essentially as previously described³³. Membranes were isolated from 5-l *E. coli* cultures and solubilized in 1% dodecyl-β-D-maltopyranoside (DDM; Genexon) for 2 h in buffer containing 1× PBS, 150 mM NaCl and 10 mM imidazole. The suspension was cleared by ultracentrifugation at 120,000g for 1 h. The sample was mixed with 1 ml of Ni-NTA Superflow resin (Qiagen) per 1 mg of GFP–His₈ and incubated for 2 h at 4 °C. Slurry was loaded onto a glass EconoColumn (Bio-Rad) and washed in 1× PBS buffer containing 0.1% DDM, 150 mM NaCl and 20 mM imidazole for 20 column volumes. Bound material was washed for a further 20 column volumes in the same buffer containing 50 mM imidazole. The NapA–GFP–His₈ fusion was eluted in two column volumes of 1× PBS buffer containing 0.6% nonyl-β-D-maltopyranoside (NM; Genexon), 150 mM NaCl and 250 mM imidazole. The eluted protein was dialysed overnight in the presence of stoichiometric amounts of His₆-tagged tobacco etch virus protease in 1.5 l of buffer containing 20 mM Tris–HCl, pH 7.5, 150 mM NaCl and 0.5% NM. Dialysed sample was passed through a 5 ml Ni-NTA His-Trap column (GE Healthcare), and the flow-through containing NapA was collected. Protein was concentrated to 10 mg ml^{−1} using concentrators with a relative molecular mass cut-off of 100 kilodaltons (kDa), and was loaded onto a Superdex 200 10/300 gel filtration column (GE Healthcare) equilibrated in 20 mM Tris–HCl, pH 7.5, 150 mM NaCl and 0.45% NM (Anatrace). The protein peak was collected and concentrated to 10 mg ml^{−1} for crystallization.

Purified NapA was loaded onto a Superdex 200 10/300 size exclusion column (GE Healthcare) coupled to a Viscotek TDAmax tetra detector array (Malvern) with GPCmax solvent pump and integrated auto-sampler, using the OmniSEC software for data analysis. The SEC–UV/LS/refractive index system was equilibrated in 20 mM Tris–HCl, pH 7.5, 150 mM NaCl and 0.03% DDM at a flow rate of 0.3 ml min^{−1}. Standard gel filtration molecular weight markers were used for calibration and all proteins were analysed under the same experimental conditions. Data were collected from the refractive index, right angle LS (RALS) and UV_{280 nm} detectors. The oligomeric state of NapA from the NapA–detergent micelle was calculated using methods described previously³⁴.

NapA transport activity. In this set-up (Fig. 1a), the described Na⁺/H⁺ antiport with inverted membrane vesicles was mimicked^{35,15}. For this, we co-reconstituted *E. coli* F₁F₀ ATP synthase and *T. thermophilus* NapA in the same vesicles. Although ATP synthase under these conditions has been shown to orient unidirectionally³⁵, NapA most probably has a heterogeneous orientation, which might affect apparent *K_m* values. Furthermore, a mixture of liposomes containing none, one or two enzymes is expected. Only the liposomes containing ATP synthase will lead to initial ACMA quenching, but only those containing both ATP synthase and NapA will lead to dequenching after addition of the coupling ion. Furthermore, it is expected that a new equilibrium is reached after every addition, which might influence the extent of dequenching. This is demonstrated by the addition of NH₄Cl at the end of every measurement, in which any remaining ΔpH is dissipated. In brief, purified NapA

wild-type and mutants were co-reconstituted with purified ATP synthase from *E. coli* with an ~2:1 molar ratio (NapA:ATP synthase) in MME buffer (10 mM MOPS–NaOH, pH 7.5, 2.5 mM MgCl₂, 100 mM KCl) as described for ATP synthase^{35,36}. Typically, 50 μl proteoliposomes were diluted into 1.5 ml MME buffer containing 3 nM ACMA and 140 nM valinomycin. Fluorescence was monitored at 480 nm using an excitation wavelength of 410 nm in a fluorescence spectrophotometer (Cary Eclipse, Agilent Technologies). An outward-directed pH gradient (acidic inside) was established by the addition of 2 mM ATP, as followed by a change in ACMA fluorescence. After ~2 min equilibration, the activity of NapA wild-type and mutants thereof was assessed by the dequenching of ACMA fluorescence after addition of the indicated concentrations of NaCl or LiCl. Addition of 20 mM NH₄Cl leads to near complete dequenching. Each experiment was performed in triplicate.

Na⁺/H⁺ antiport activity by NapA is electrogenic. In this set-up, we followed ΔpNa or Δψ driven H⁺ transport as a consequence of electrogenic Na⁺/H⁺ exchange activity. In this sensitive assay, liposomes containing the highly soluble and membrane impermeable pH sensitive dye pyranine were used to follow H⁺ influx or efflux. A Na⁺ gradient was established by the addition of NaCl, whereas an electrical membrane potential was established with a K⁺/valinomycin diffusion potential.

Reconstitution of NapA into liposomes containing pyranine was essentially performed as described³⁷. In brief, to a 500 μl liposome (40–80 nm) suspension (soy bean lipids, type II, SIGMA, 20 mg ml^{−1}) in buffer A (10 mM MOPS–PO₄, pH 7.5), 45 μl cholate (20% stock solution) and 17 μl NapA (50 μM, purification buffer) was added and incubated for 30 min at room temperature with occasional mild mixing. The cholate was removed via a PD-10 gel filtration column (GE Healthcare) equilibrated with buffer A and the proteoliposomes in the void volume were collected (~1.2 ml). They were diluted to 8 ml with buffer A, collected via ultracentrifugation (200,000g, 4 °C, 30 min) and resuspended in 250 μl buffer A. Then, 125 μl of proteoliposomes was mixed with 1 mM pyranine (0.1 M stock solution) and the desired Na₂SO₄ and K₂SO₄ concentrations, frozen in liquid nitrogen, thawed in water and briefly sonicated in a bath type sonicator (2 × 5 s). The freeze/sonication procedure was repeated once. The external pyranine was subsequently removed via a prepacked G25 gel filtration column (GE Healthcare) and the proteoliposomes, equilibrated in buffer A, were collected from the void volume of the column.

Pyranine fluorescence measurements monitoring pH changes on the inside of the proteoliposomes were performed as described³⁷. Typically, an amount of 20 μl liposomes containing the desired Na⁺ and K⁺ concentrations was mixed with 2.5 ml buffer containing the same buffer with the appropriate salt concentrations. After 30 s, exchange activity was either initiated by the addition of 50 mM Na⁺ to the outside (0.5 mM Na⁺ on the inside), establishing an inwardly-directed Na⁺ gradient (~120 mV), and leading to H⁺ efflux. Accordingly, in a system in the absence of Na⁺ gradient (50 mM Na⁺ on both sides), but in the presence of a K⁺ gradient (100 mM K⁺ inside, 1 mM K⁺ outside liposomes), addition of valinomycin (10 nM) established a membrane potential of ~120 mV (inside negative) driving H⁺ influx (and Na⁺ efflux).

Crystallization and preliminary screening. Crystals were grown at 20 °C using the hanging drop vapour diffusion method. A 1 μl aliquot of pure protein was mixed 1:1 with reservoir solution containing 0.001 M zinc sulphate, 0.05 M HEPES, pH 7.8 and 22–36% PEG 600 (mutants) or 0.05 M glycine, pH 9.0, 0.05 M magnesium acetate, 26% PEG 400 (wild type). Crystals appeared overnight and reached maximum size after 3–4 days. For specified crystals, coverslips were transferred for overnight incubation with reservoir solution containing 2% increments of PEG 400. Dehydrated crystals in 30–36% PEG 400 were finally soaked with 1 μl reservoir solution containing 1% NM and 40% PEG 400 followed by flash freezing in liquid nitrogen before data collection.

Data were collected at the European Synchrotron Radiation Facility and Diamond Light Source. Most of the crystals were triclinic, however, very occasionally crystals with different space groups were observed. The highest resolution data were collected from an orthorhombic crystal of the triple mutant that had been reannealed on the beamline. This crystal was grown with 0.025% dichloromethane (Hampton Research) as an additive.

Structure determination. To obtain phases cysteine mutants were introduced into the protein to enable derivatization with mercury. Three positions were chosen (Met 20, Val 166 and Val 326) and single, double and triple mutants created as described above. Mercury-derivatized protein was prepared by incubation of the protein at 20 °C for 1 h with 2.5 mM mercury acetate. The structure was solved using MIRAS from four triclinic crystals (native, mercury derivatized and selenomethionine) as shown in Supplementary Table 1. Data were processed using the Xia2 pipeline³⁸ to XDS³⁹, with further processing using the CCP4 suite of programs⁴⁰. Heavy atom sites were located from anomalous difference Patterson maps of the double mutant using the program RSPS⁴¹. Phases were calculated and refined

in SHARP⁴². Further mercury and selenomethionine sites were located in residual maps. The crystals contain four molecules in the asymmetric unit. Non-crystallographic symmetry operators were determined from the heavy atom positions and averaging was carried out in DM⁴³ with a mask calculated in O⁴⁴. This map was then used as a search model for molecular replacement in Phaser⁴⁵ for the P₂₁ and C222₁ data sets. On the basis of the operators obtained, multi-crystal averaging in DMULTI⁴³ using the P₁ (4 molecules in au), P₂₁ (2 molecules in au) and C222₁ (1 molecule in au) data sets was carried out. This gave maps of sufficient quality to see all 13 helices of the NapA subunit. Model building was carried out in O⁴⁴ and COOT⁴⁶ and was facilitated by the positions of the mercury and selenomethionine peaks. Refinement of the atomic coordinates and individual B-factors using the C222₁ data at 2.9 Å was carried out in PHENIX⁴⁷. Secondary structure restraints were applied and refinement was interspersed with rebuilding using O and COOT. Peaks in anomalous difference maps indicated the presence of two zinc ions bound to the protein on the periplasmic surface to residues at the N terminus of TM-1 and the loop between TM1 and TM2. The final refinement statistics are shown in Supplementary Table 1.

Maps were calculated based on the final refined model for the data from the wild type selenomethionine derivatized crystals (Supplementary Table 1). At 3.7 Å resolution no clear differences were observed and the positions of the peaks in the anomalous difference maps were consistent with the positions of the methionines in the structure (data not shown).

Superpositions were carried out in Lsqman⁴⁸ such that all matching C α pairs were less than 3.8 Å apart after superposition. Figures were drawn using Pymol⁴⁹ except those showing electron density, which were made using CCP4mg⁵⁰. The inward-facing model of NapA was created by superposing the core and dimerization domains of NapA onto the corresponding domains of NhaA separately. The only adjustments that were made to the model were to the polypeptide chain between the two domains. No modifications were made to alleviate the clashes between the loop between TM-1 and TM1 and TM4a. The video was made in Lsqman⁴⁸ by morphing between the outward-facing crystal structure and the inward-facing model displayed using Pymol⁴⁸.

Molecular dynamics simulations. Molecular dynamics simulations of the NapA dimer in a mixed POPE/POPG bilayer were carried out with the Gromacs simulation package⁵¹, either version 4.5.5 or a development version of 4.6. All simulations employed the CHARMM force field including CMAP^{52,53} with the original TIP3P water model⁵⁴ and updated CHARMM parameters for POPE and POPG lipids (CHARMM)⁵⁵, as implemented in Gromacs⁵¹. The ratio of POPE to POPG molecules was about 4:1 to approximate the major components of the *E. coli* membrane. We used a multi-scale approach to embed the dimer into the membrane⁵⁶. The protein was first simulated in a coarse grained representation⁵⁷ and the membrane was allowed to self-assemble around the protein from a random mixture of lipids and water in the simulation box⁵⁸. After 200 ns simulation with an integration time step of 20 fs the bilayer had assembled around the NapA dimer. In the second step of the multi-scale approach, the system was converted to the CHARMM atomistic representation with the CG2AT protocol⁵⁶ and the original crystal structure inserted in place of the back-translated protein. The simulation system consisted of an orthorhombic simulation box of size 114 Å \times 114 Å \times 91 Å containing 118,832 atoms in 768 protein residues, 215 POPC and 57 POPG lipids, 180 Na⁺, 109 Cl⁻ ions and 24,575 water molecules. The free NaCl concentration was about 250 mM in all simulations to approximate the outward-open transporter facing an environment of increased salt stress.

Most titratable residues were predicted by PROPKA⁵⁹ to be in their default charge states at pH 7.8 and were simulated as such (including deprotonated Asp 156 and Asp 157, with predicted pK_a as 3.2 and 5.7). The pK_a of the buried Lys 305 was predicted to be 9.5, with its charged form partially stabilized by a salt-bridge with the (charged) Asp 156. Therefore, Lys 305 was simulated in its protonated (positively charged) form. For a number of residues the predicted value was within 1 unit of the environmental pH value we wished to simulate and for those residues we rationalized our choices as follows: The pK_a of Glu 333 was predicted as 7.0. Glu 333 might help to stabilize the helix dipoles of helices TM4a and TM11a (similar to Asp 133 in NhaA) but such a charge-dipole interaction is not encoded in the empirical rules of the PROPKA algorithm⁵⁹. Furthermore, simulations clearly showed that Glu 333 protrudes into the outward facing cavity and is solvated. We, therefore, adopted a charged Glu 333. His 6 and His 51, which reside on the surface and do not seem to have any specific mechanistic role, had predicted pK_a of 7.8 and 7.1 and were modelled in their dissociated (neutral) form.

Equilibrium molecular dynamics simulations were performed with periodic boundary conditions at constant temperature $T = 323$ K and pressure $P = 1 \times 10^5$ Pa using the velocity rescaling algorithm for the thermostat (time constant 0.1 ps)⁶⁰ and semi-isotropic Parrinello–Rahman barostat (time constant 5.0 ps, compressibility 4.6×10^{-10} Pa⁻¹, coupling every 10 steps)⁶¹. Long range corrections for energy and pressure were applied⁶². Lennard–Jones interactions were switched off between 8 Å

and 12 Å, and electrostatic interactions were handled by the SPME method⁶³ in which Coulomb interactions were computed in real space up to a cut-off of 12 Å and long range interactions beyond the cut-off were calculated in reciprocal space with fast Fourier transforms on a grid with spacing 1.2 Å and fourth order splines for fitting of the charge density. The grid-based neighbour list was updated every five steps to a distance of 14 Å. Bonds to hydrogen atoms were constrained with the P-LINCS algorithm⁶² or SETTLE (for water molecules)⁶⁴. The classical equations of motions were integrated with a leap frog integrator and a time step of 2 fs. Conformations were saved every 1 ps for analysis.

The simulation protocol included an initial energy minimisation of the atomistic system and a 1-ns equilibrium simulation during which the protein heavy atoms were restrained with a harmonic force with force constant of 1,000 kJ mol⁻¹ nm⁻². An initial unrestrained simulation of the dimer with Asp 156 and Asp 157 in their default protonation state was run for 100 ns. Eight additional 100-ns simulations were performed in three sets (two simulations in set 1, five in set 2, one in set 3). The starting configuration for each set was generated from the last frame of the initial simulation by exchanging any sodium ion within 3 Å of Asp 157 or Asp 156 with a random bulk water molecule. Repeats in each set always differed by the seed of the random number generator, thus leading to differing initial assignments of velocities and generation of independent trajectories through the stochastic component of the velocity rescaling thermostat⁶⁰, as seen from the different r.m.s.d. time series in Supplementary Fig. 6.

Additional simulations were performed to assess the influence of the protonation state of the conserved residues Asp 156 and Asp 157 on sodium binding as performed previously for NhaA⁸. Two independent 100-ns simulations were performed for each of (1) Asp 156 deprotonated (negatively charged) and Asp 157 protonated (neutral), (2) Asp 156 protonated and Asp 157 deprotonated, and (3) both Asp 156 and Asp 157 protonated. Charge states were modified with the Gromacs tool `pdb2gm`⁵⁰. These six simulations used the 100-ns frame of the initial simulation (default charge states) with any sodium ions near the aspartates exchanged with a bulk water molecule as a well-equilibrated starting conformation. In each case, a 1-ns position restraint simulation (as above) was followed by a 100-ns production equilibrium simulation.

Simulations were analysed with MDAnalysis⁶⁵ and Gromacs tools⁵¹. To calculate the sodium density, data from all nine simulations with the deprotonated aspartates were used at 1-ps intervals. Ion binding and unbinding to each protomer appeared to be independent so that data for both protomers were combined by superpositioning trajectories of both protomer A and protomer B on the coordinates of protomer A from the start of the initial simulation. The density was calculated by histogramming sodium coordinates in cubic volume elements at a resolution of 1 Å in a fixed coordinate system defined by the initial coordinates of protomer A. The thicknesses of bilayer regions were calculated from the distributions of the headgroup phosphate and acyl chain atoms along the membrane normal, using trajectories superpositioned on the dimer.

Images showing simulation data were prepared with VMD⁶⁶ and the Bendix plugin for curved helices⁶⁷ or UCSF Chimera^{68,69}.

29. Drew, D. E., von Heijne, G., Nordlund, P. & de Gier, J. W. Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*. *FEBS Lett.* **507**, 220–224 (2001).
30. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
31. Wagner, S. et al. Tuning *Escherichia coli* for membrane protein overexpression. *Proc. Natl Acad. Sci. USA* **105**, 14371–14376 (2008).
32. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
33. Drew, D. et al. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nature Protocols* **3**, 784–798 (2008).
34. Slotboom, D. J., Duurkens, R. H., Olieman, K. & Erkens, G. B. Static light scattering to characterize membrane proteins in detergent solution. *Methods* **46**, 73–82 (2008).
35. Wiedenmann, A., Dimroth, P. & von Ballmoos, C. Functional asymmetry of the F₀ motor in bacterial ATP synthases. *Mol. Microbiol.* **72**, 479–490 (2009).
36. Ishmukhametov, R. R., Galkin, M. A. & Vik, S. B. Ultrafast purification and reconstitution of His-tagged cysteine-less *Escherichia coli* F₁F₀ ATP synthase. *Biochim. Biophys. Acta* **1706**, 110–116 (2005).
37. Wiedenmann, A., Dimroth, P. & von Ballmoos, C. $\Delta\psi$ and Δ pH are equivalent driving forces for proton transport through isolated F₀ complexes of ATP synthases. *Biochim. Biophys. Acta* **1777**, 1301–1310 (2008).
38. Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
39. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
40. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
41. Knight, S. D. RSPS version 4.0: a semi-interactive vector-search program for solving heavy-atom derivatives. *Acta Crystallogr. D* **56**, 42–47 (2000).

42. Bricogne, G., Vornrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
43. Cowtan, K. 'dm': An automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newslett. Prot. Crystallogr.* **31**, 34–38 (1994).
44. Jones, T. A., Kjeldgaard, M., Charles, W. C. Jr & Robert, M. S. Electron-density map interpretation. *Methods Enzymol.* **277**, 173–208 (1997).
45. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
46. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
47. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
48. Kleywegt, G. J. & Jones, T. A. A super position. *Joint CCP4 ESF-EACBM Newslett. Prot. Crystallogr.* **31**, 9–14 (1994).
49. Delano, W. L. The PyMOL Molecular Graphics System (DeLano Scientific, 2002).
50. Potterton, L. *et al.* Developments in the CCP4 molecular-graphics project. *Acta Crystallogr. D* **60**, 2288–2294 (2004).
51. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
52. Mackerell, A. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).
53. Mackerell, J. A. D. Jr, Feig, M. & Brooks, C. L. III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 (2004).
54. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
55. Klauda, J. B. *et al.* Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).
56. Stansfeld, P. J. & Sansom, M. S. P. From coarse grained to atomistic: a serial multiscale approach to membrane protein simulations. *J. Chem. Theory Comput.* **7**, 1157–1166 (2011).
57. Bond, P. J., Wee, C. L. & Sansom, M. S. P. Coarse-grained molecular dynamics simulations of the energetics of helix insertion into a lipid bilayer. *Biochemistry* **47**, 11321–11331 (2008).
58. Scott, K. A. *et al.* Coarse-grained MD simulations of membrane protein-bilayer self-assembly. *Structure* **16**, 621–630 (2008).
59. Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **61**, 704–721 (2005).
60. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
61. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
62. Hess, B. P-LINCS: a parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **4**, 116–122 (2008).
63. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8592 (1995).
64. Miyamoto, S. & Kollman, P. A. SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comput. Chem.* **13**, 952–962 (1992).
65. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).
66. Humphrey, W., Dalke, A. & Schulten, K. VMD—visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
67. Dahl, A. C. E., Chavent, M. & Sansom, M. S. P. Bendix: Intuitive helix geometry analysis and abstraction for VMD. *Bioinformatics* **28**, 2193–2194 (2012).
68. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
69. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).

CORRIGENDUM

doi:10.1038/nature12528

Corrigendum: Weakened stratospheric quasibiennial oscillation driven by increased tropical mean upwelling

Yoshio Kawatani & Kevin Hamilton

Nature **497**, 478–481 (2013); doi:10.1038/nature12140

In this Letter, the QBO amplitudes were calculated incorrectly. The corrected values (consistent with the description of the methods in the paper) are all larger than those reported, generally by about 10–20%. The error has only a very small effect on the percentage trends of any of the time series examined, and the conclusions of the original paper are unaffected. The detailed claims throughout are unaffected, except that the significance threshold for the QBO amplitude trend reported for the MIROC-ESM-CHEM simulation during the 1953–2012 period should be $P < 0.08$ instead of $P < 0.06$ as reported originally. This error and Figs 1–3 and Supplementary Figs 3 and 9 have been corrected in the HTML and PDF versions of the original paper. We thank L. D. Oman for bringing this to our attention.

CORRIGENDUM

doi:10.1038/nature12641

Corrigendum: LRG1 promotes angiogenesis by modulating endothelial TGF- β signalling

Xiaomeng Wang, Sabu Abraham, Jenny A. G. McKenzie, Natasha Jeffs, Matthew Swire, Vineeta B. Tripathi, Ulrich F. O. Luhmann, Clemens A. K. Lange, Zhenhua Zhai, Helen M. Arthur, James W. B. Bainbridge, Stephen E. Moss & John Greenwood

Nature **499**, 306–311 (2013); doi:10.1038/nature12345

In Supplementary Fig. 13 of this Article, we inadvertently duplicated two of the representative retinal vascular flat-mount images. The wild-type P10 superficial plexus was duplicated and shifted approximately 10 degrees and displayed as the *Lrg1*^{-/-} P10 superficial plexus. The wild-type P25 deep plexus was duplicated and inverted and displayed as the *Lrg1*^{-/-} P25 deep plexus. The Supplementary Information of this Corrigendum shows the corrected Supplementary Fig. 13, in which the duplicated images have been replaced with representative images of the *Lrg1*^{-/-} mouse P10 superficial plexus and the *Lrg1*^{-/-} mouse P25 deep plexus. We have reanalysed all of the correct images and find our conclusions unchanged.

Supplementary Information is available in the online version of this Corrigendum.

ERRATUM

doi:10.1038/nature12619

Erratum: Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage

Mohamed F. Haroon, Shihu Hu, Ying Shi, Michael Imelfort, Jurg Keller, Philip Hugenholtz, Zhiguo Yuan & Gene W. Tyson

Nature **500**, 567–570 (2013); doi:10.1038/nature12375

In this Letter, equation (1) was inadvertently shown incorrectly, with CO₂ missing from the reaction products. The correct equation (1) is shown below:



This has been corrected in the HTML and PDF versions of the original manuscript.

CAREERS

NATUREJOBS BLOG Expert careers advice, updates and discussions go.nature.com/z8g4a7

@NATUREJOBS Follow us on Twitter for the latest news and features go.nature.com/e492gf

NATUREJOBS For the latest career listings and advice www.naturejobs.com

ADAM NIKLEWICZ/IMAGES.COM/CORBIS



BY GENE RUSSO

After finishing her PhD in April, Kristina Stemler came to a crossroads. Should she pursue the tantalizing dream of pure research in academia, or find an industry job in which product sales govern priorities and shape career prospects? Her decision depended on many considerations — what did she enjoy most? What was most satisfying? Where would the hottest jobs be? And then there was that other matter looming in the back of her mind and sometimes keeping her up at night: the fact that she is tens of thousands of dollars in debt.

Stemler is one of many junior researchers who face huge debt on top of the already-daunting burdens of getting funding, landing a position and toiling for long hours in the lab. Debt can cause stress and change the calculus of career decisions. It does not usually accumulate as a result of graduate studies — for which students often get fellowships or other funding — but from the ever-increasing cost of an undergraduate education. In the United States, tuition fees have increased faster than inflation, and total student debt now exceeds the nation's credit-card debt.

Managing the bills is an ever-growing challenge. But even if early-career scientists have the resources to avoid amassing huge debt, they still have to face fiscal realities. Graduate students must learn to get by on sometimes meagre stipends while still planning ahead for big costs — buying a home, setting aside a rainy-day fund or squirrelling away cash for retirement. In a tight funding climate and tough academic job market, they are discovering the importance of financial planning.

DEBT COLLECTION

Gauging the scope of the debt problem for graduate students is difficult, but the 2013 Nature Careers Graduate Student Survey included questions intended to measure money anxiety (for further results, see 'Graduate trends'). Of more than 5,000 graduate students who responded by e-mail, on websites and through social media, 39% were 'concerned' or 'very concerned' about debt. (Even more may have been so concerned about debt that they left higher education, but the survey reached only current students). According to the US National Science Foundation, life scientists who earned a PhD in the United States in 2011 held an average debt of US\$17,426; physical scientists had an average of \$12,146.

For graduate students around the world, ►

EDUCATION

Financial burden

Graduate students face big decisions about money. They can benefit from wise counsel and careful forethought.

► debt can be a career game-changer (see ‘Money woes in China’). Dismayed by her financial burden and the bleak job prospects in academia, Stemler is planning to move into translational medicine and then, she hopes, into cancer-diagnostics research at a biotechnology or pharmaceutical company. To help with that, she is about to embark on a postdoc at the University of Texas MD Anderson Cancer Center in Houston, where she wants to focus on the mitigation of chemotherapy symptoms and how cells respond to treatment. “I was already disillusioned with the situation now in academia,” she tells *Nature* while on a break from driving from her graduate school, Washington University in St. Louis, Missouri, to Houston. “And then I had financial worries — and there was the sequester” (see *Nature* 498, 527–528; 2013).

“In a tight funding climate and tough academic job market, students are discovering the importance of financial planning.”

But some students, such as survey respondent Aditi Dubey, are still intent on pursuing academia. Dubey, who is undertaking a molecular-biology PhD at Rutgers University in New Brunswick, New Jersey, was fortunate in that her parents helped her to pay for her undergraduate degree at Drexel University in Philadelphia, Pennsylvania. She also took a part-time job. Now in the fifth year of her PhD, she is relatively debt-free; her annual \$28,000 stipend has been sufficient, although she does have to miss the occasional conference if her principal investigator does not have the funding. But Dubey is cognizant of her financial constraints, and she covets the

milestone acquisitions of many young professionals: a car, a house.

Dubey plans to do a postdoc in neuroscience, and to stay in academia unless she fails to secure a professorship after a sustained, concerted effort. She revels in the freedom, pace and culture of academic research — benefits that she found lacking on a research placement at pharmaceutical company Merck in West Point, Pennsylvania, during her undergraduate studies. Industry is Dubey’s plan B; her plan C is to become a consultant on life-sciences research. “Finances could be why I give up on the academic track,” she says.

Michael Sunde, who studies geographic information systems (GIS) at the University of Missouri in Columbia, has similar financial concerns. His skills put him in demand in the private sector, but he remains keen on academia, aiming for a post in which he can apply GIS to urban-growth issues such as modelling and tracking rainwater runoff. Private-sector opportunities are more plentiful and, in many cases, better paid — he has several classmates who have been recruited by defence contractors interested in monitoring objects of interest, not weather. But academia remains Sunde’s ideal despite his debt of about \$50,000. “There’s so much anxiety that I can’t think about it,” he says. “I have to push it aside.”

FINANCIAL MANAGEMENT

The US Council of Graduate Schools (CGS) in Washington DC, an association of US universities, has taken notice of such cases. Last November, the CGS and TIAA-CREF, a financial-services provider based in Charlotte, North Carolina, launched the first phase of a three-year programme to improve the debt-management and financial-planning abilities of graduate and

undergraduate students. So far, 15 institutions have been selected to participate; each received grants to develop financial-literacy projects, assess students’ needs and augment financial-planning resources and tools. Another 19 affiliate university partners will be invited to share education tools and promising practices, and to take part in the programme’s study of graduate students’ financial habits.

The impetus for the project involved a number of factors, says Daniel Denecke, a CGS associate vice-president who is heading the effort. “It was the rising amount of student debt we were seeing nationally: how hardship and issues are increasingly playing a role in student decisions to defer plans to go into grad school.” Programme leaders noted that student borrowing is not necessarily informed by longer-term salary potential, and wanted to help students “to make informed decisions when they do borrow”, says Denecke, “by having access to projections on how the amount they borrow will affect their career decisions”.

The Ohio State University (OSU) in Columbus is among the programme’s grantees. The university’s graduate school and its Student Wellness Center are planning online tutorials that will offer guidance on such tasks as compiling a monthly budget or assessing the riskiness of financial practices. They will send out financial surveys, and counsellors will use the responses to reach out to students who seem to be at high risk of getting into difficulties. Part of the OSU’s strategy is to tailor its advice to key phases in student life — the admissions process, candidacy after completing initial course requirements and graduation, says graduate-school dean Patrick Osmer. For example, newly admitted students probably do not need information about retirement, but they should have insight into the salary prospects in a given field.

Osmer also supports training graduate students to be peer counsellors, reasoning that students respond better when they hear something from their peers rather than from an administrator or instructor.

CLEAN CREDIT

Cornell University in Ithaca, New York — another CGS grant recipient — is planning to pool the resources of multiple campus offices to educate students about debt and financial management both in person and online. The university already works with ClearPoint, a financial-advice non-profit based in Richmond, Virginia, which conducts workshops and telephone consultations focused on budgeting, financial management and handling debt.

The workshops have been very well attended, says Gary Thurber, who was until recently a ClearPoint credit counsellor and educator. He had expected students to be most interested in budgeting and goal-setting — working out, for example, how to make a stipend instalment received in August last until January. But what most students want to know about, he says,

GLOBAL OUTLOOK

Money woes in China

Researchers in China face some of the most acute debt worries, according to the 2013 Nature Careers Graduate Student Survey. In China, 67% of respondents were ‘concerned’ or ‘very concerned’ about their debt. Of the other countries with sizeable respondent samples, the Netherlands had the fewest expressing concern, at 17%; the United Kingdom was next lowest at 20%. In the United States, 26% of respondents were concerned or very concerned.

For many students in China, and Asia in general, such worries are attributable to rising education costs and tough job markets, says Denis Simon, an expert on China and vice-provost for international strategic initiatives at Arizona State University in Tempe. Increasingly, he notes, Chinese students who come to the United States for what they

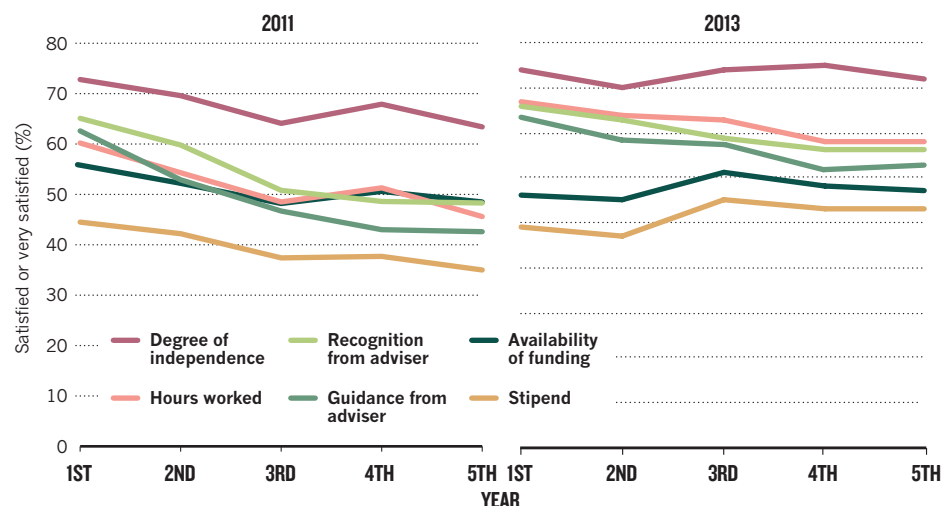
hope will be a superior education are not just the wealthy, but also from the middle class. Their families struggle to pay the US\$40,000 per year or more for a full undergraduate education — much more than in China. There is little financial assistance available for foreign students in the United States, and it is hard for them to get jobs off campus for more than 10–15 hours a week. Those who go on to graduate studies hold large debt — and many graduates, says Simon, have difficulty finding well-paid jobs. The costs of living in Beijing or Shanghai are challenging, whether one has studied abroad or not.

“The concern about having to be successful upon graduation runs deep,” says Simon, “especially for Chinese, who may borrow from within their families and relatives.” **G.R.**

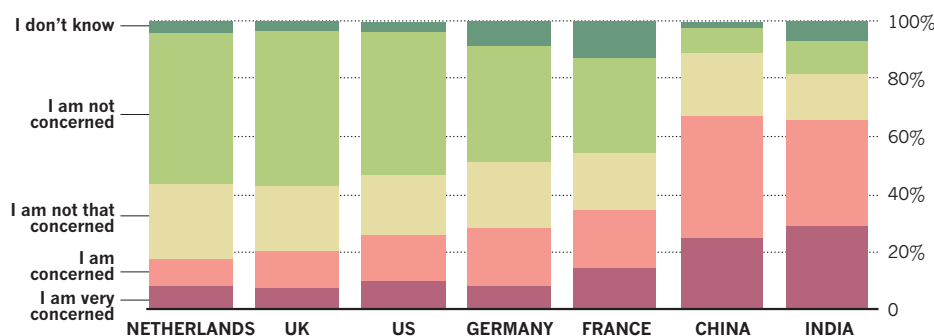
Graduate trends

The 2013 Nature Careers Graduate Student Survey collected more than 5,000 responses from graduate students around the world. The questions, like those of the inaugural 2011 survey, probed concerns about funding, working conditions and the factors affecting careers paths; this year's survey also asked about student debt.

Q To what extent are you satisfied with each of the following attributes or aspects of your graduate-school experience?



Q How concerned, if at all, are you about any debt you will incur from pursuing your post-graduate qualification or degree?



Q Which, if any, of the following are reasons that you feel less likely to follow a traditional research career than you did when you started your PhD?



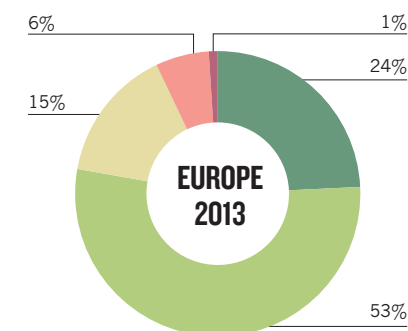
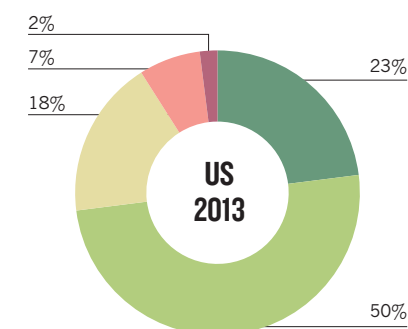
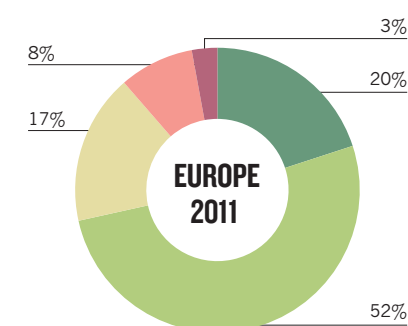
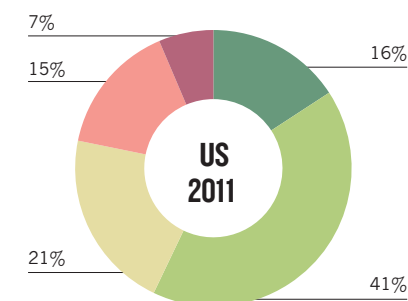
are credit reports and the best ways to obtain mortgages. He tells them how to get free credit reports (avoid the unnecessary charges of the oft-advertised freecreditreport.com, and opt instead for the federally authorized www.annualcreditreport.com, which delivers one free report per year from each of the three US credit-reporting agencies). He also advises

students to keep their credit-card balances below 40% of the credit limit, and to be wary of transferring balances between cards.

Still, Thurber does discuss debt, advising students to borrow only what they need. Even with low interest rates, payments add up, he notes. He often counsels students to get a part-time job to cover basic expenses.

Q Overall, how satisfied are you with your graduate-school experience? By region.*

Very satisfied Satisfied Very dissatisfied
Neutral Dissatisfied



*Owing to rounding, numbers may not add up to 100%

"In our colleges and schools," says Thurber, "we do a great job of educating our people for jobs they're trained to do. But I don't think right now that many colleges are teaching students about personal finance. We're not teaching them what to do with income once they have it." ■

Gene Russo is *Careers* editor at Nature.

THE RUMINATION ON WHAT ISN'T

The right move.

BY ALEX SHVARTSMAN

This isn't a time-travel story. As you sit in the sterile room that looks and feels like a hospital ward but doesn't smell or sound like one — you know the difference after so many months — as you stare at the thing at the other end of the chessboard and try to picture your daughter instead, you ponder the choices you've made. You realize that, if given a chance, you wouldn't go back to change anything. There is no decision, no single action that would have altered the course of events. No moment in time that could have been modified to avoid this outcome. No matter what, you would find yourself in this room, lost in thought as your fingers caress a white pawn. She is waiting for you to make your move.

This isn't a horror story. You certainly felt like it was, on that day in the doctor's office, when he delivered the diagnosis with practised compassion. Words like *stage four* and *metastasis* sounded surreal. They were fears of old people; there was no place for them in the life of a ten-year-old girl. You remember feeling shaken and detached, as if this was happening to somebody else. But the oncologist wouldn't allow you to process this fully, to despair, to grieve. He wanted to talk treatment options, and DNA sequencing, and clinical trials. And he wanted decisions to be made right away, because there wasn't a lot of time.

But it wasn't all fear and dread. There were moments of happiness, when the two of you giggled while watching cartoons together, or went picking apples under the pleasantly warm September sun. There were moments of boredom, hours spent in waiting rooms filled with year-old magazines and stone-faced strangers who probably understood what you were going through

better than your closest friends, but were barricaded behind the walls of their own distress. And then there were mundane moments, because even when your world is shattered you still have to go through the motions of picking up paper towels at the supermarket, and getting the oil changed in your truck, and doing laundry.

This isn't a fantasy story. When her hair fell out from the chemo, and she lost weight, and the doctors began to mention the word hospice, you sought alternative treatments. Folk medicines and psychics, and any number of other things that don't work, but desperate people try them anyway, because they're better than doing nothing. This isn't a fantasy, there's no magic or miracles. She kept getting worse.

This isn't exactly a tragedy. Before the illness could take her, lawyers in expensive suits showed up and offered you a way out. It was an experiment,

they said. The procedure had never been performed on a human being before. There were plenty of risks and unknowns, but your daughter was given an opportunity to make history. More importantly, it was the way for her to survive. You hated the thought of her being used as some sort of a guinea pig, but the doctors said that she had just a few weeks left. This was her only chance. So you signed endless pages of legal documents until your hand ached almost as much as your heart, and you allowed yourself to hope.

This is a love story. You sit in the sterile room and play chess against a sleek metal box that houses your daughter's mind. She has been uploaded, the first herald of the coming singularity. She will never again pick apples, or play soccer, or hug you. The scientists don't know whether her mind will continue to develop or if she will remain a perpetual ten-year-old. They don't know whether she will live forever inside the machine, or if her consciousness will degrade and disappear with time. The entire world is waiting to find out.

She watches you through digital cameras and hums her favourite tune through the speakers, impatient for you to finally make your move. And you know with absolute certainty that this is a love story, because you love your daughter just the same, regardless of her physical form. You get to tell her stories, and watch cartoons together, and play chess, and face whatever challenges may come in the future as a family.

You smile at her, and push the pawn forward. ■

Alex Shvartsman is a writer and game designer from Brooklyn, New York. His other fiction is linked at www.alexshvartsman.com.

JACEY

➤ **NATURE.COM**
Follow Futures:
@NatureFutures
go.nature.com/mtoodm

Low-voltage magnetoresistance in silicon

ARISING FROM C. H. Wan, X. Z. Zhang, X. L. Gao, J. M. Wang & X. Y. Tan *Nature* **477**, 304–307 (2011)

Magnetoresistance exhibited by non-magnetic semiconductors has attracted much attention^{1–13}. In particular, Wan *et al.* reported room-temperature magnetoresistance in silicon to reach 10% at 0.07 T and 150,000% at 7 T—“an intrinsically spatial effect”¹². Their supply voltage was approximately 10 V (ref. 12), which is low and approaches the industrial requirement¹⁴. However, we have found their large magnetoresistance values to be experimental artefacts caused by their method of measurement. The true room-temperature magnetoresistance of the devices described in ref. 12 is low with a magnetic field of up to 7 T and a supply voltage of around 10 V and hence these devices cannot offer large magnetoresistance with low supply voltage to industry. There is a Reply to this Brief Communication Arising by Zhang, X. Z., Wan, C. H., Gao, X. L., Wang, J. M. & Tan, X. Y. *Nature* **501**, <http://dx.doi.org/10.1038/nature12590> (2013).

Wan *et al.*¹² measured two types of In/SiO₂/Si/SiO₂/In devices using a Keithley 2400 sourcemeter as both a current source and a voltage meter (which we refer to here as method 1), and obtained large magnetoresistance values of up to 10% at 0.07 T and 150,000% at 7 T. We fabricated two devices with the same structures as those of ref. 12 and performed method 1 using them. Their voltage–current (*V*–*I*) curves can be divided into different regions with different resistances, just as in the results of ref. 12. Wan *et al.*¹² claim that injection of minority carriers into silicon causes a p–n junction and the changes in resistance, that large magnetoresistance occurs with applied current in one of the regions (referred as to the transition region), and that the magnetic-field dependence of the magnetoresistance in the transition region is different from those in the other regions. However, when we used another method (here called method 2) with unchanged measuring parameters and different instruments on the devices, the *V*–*I* characteristics without the transition region were obtained. The only difference between the two methods is that in method 2 we used the Keithley 2400 only as the current source, with an independent voltmeter (Keithley 2182) as the voltage meter.

Further, we performed both methods on two circuits composed of linear resistors, which were used to simulate the devices. The results indicate that in method 1 the Keithley 2400 itself interferes with the measurement of the specimen and cannot give correct voltage values when the applied current exceeds a certain value and falls in the transition region. Because ref. 12 claims that large magnetoresistances were measured when *I* was in the transition region, magnetoresistance was defined as $[R(B) - R(B = 0)]/R(B = 0)$ and $R = V/I$, we conclude that the large magnetoresistance values are really experimental artefacts caused by the interference of the sourcemeter. Method 2 is valid. Using it, we obtained magnetoresistance values for the two devices with supply voltages of 6.7–72 V and 0.79–50 V, respectively. The values are all low and the magnetic-field dependence at all applied

currents is the same (above 2 T the field dependence is linear); the magnetoresistance does not exhibit any signs of saturation at fields up to 7 T. The linear dependence without magnetoresistance saturation is the same as for inhomogeneity-induced magnetoresistance^{7–9}.

Jun Luo¹, Peisen Li², Sen Zhang², Hongyu Sun¹, Hongping Yang¹ & Yonggang Zhao²

¹Beijing National Center for Electron Microscopy, Laboratory of Advanced Materials, Department of Materials Science and Engineering, Tsinghua University, Beijing 100084, China.

email: jluc@mails.tsinghua.edu.cn

²Department of Physics and State Key Laboratory of Low-Dimensional Quantum Physics, Tsinghua University, Beijing 100084, China.

Received 20 September 2012; accepted 16 August 2013.

1. Smith, R. A. *Semiconductors* 2nd edn, Ch. 5 (Cambridge Univ. Press, 1978).
2. Popović, R. S. *Hall Effect Devices* Ch. 3 (IOP Publishing, 1991).
3. Xu, R. *et al.* Large magnetoresistance in non-magnetic silver chalcogenides. *Nature* **390**, 57–60 (1997).
4. Solin, S. A., Thio, T., Hines, D. R. & Heremans, J. J. Enhanced room-temperature geometric magnetoresistance in inhomogeneous narrow-gap semiconductors. *Science* **289**, 1530–1532 (2000).
5. Lee, M., Rosenbaum, T. F., Saboungi, M.-L. & Schnyders, H. S. Band-gap tuning and linear magnetoresistance in the silver chalcogenides. *Phys. Rev. Lett.* **88**, 066602 (2002).
6. Husmann, A. *et al.* Megagauss sensors. *Nature* **417**, 421–424 (2002).
7. Parish, M. M. & Littlewood, P. B. Non-saturating magnetoresistance in heavily disordered semiconductors. *Nature* **426**, 162–165 (2003).
8. Hu, J. S. & Rosenbaum, T. F. Classical and quantum routes to linear magnetoresistance. *Nature Mater.* **7**, 697–700 (2008).
9. Delmo, M. P., Yamamoto, S., Kasai, S., Ono, T. & Kobayashi, K. Large positive magnetoresistive effect in silicon induced by the space-charge effect. *Nature* **457**, 1112–1115 (2009).
10. Delmo, M. P., Kasai, S., Kobayashi, K. & Ono, T. Current-controlled magnetoresistance in silicon in non-Ohmic transport regimes. *Appl. Phys. Lett.* **95**, 132106 (2009).
11. Schoonus, J. J. H. M., Haazen, P. P. J., Swagten, H. J. M. & Koopmans, B. Unravelling the mechanism of large room-temperature magnetoresistance in silicon. *J. Phys. D* **42**, 185011 (2009).
12. Wan, C. H., Zhang, X. Z., Gao, X. L., Wang, J. M. & Tan, X. Y. Geometrical enhancement of low-field magnetoresistance in silicon. *Nature* **477**, 304–307 (2011).
13. Porter, N. A. & Marrows, C. H. Linear magnetoresistance in *n*-type silicon due to doping density fluctuations. *Sci. Rep.* **2**, 565 (2012).
14. Allan, A. *et al.* 2011 edn, <http://www.itrs.net/Links/2011ITRS/Home2011.htm> (2011).

Author Contributions J.L. designed the research blueprint, performed the experiments and data analysis, and wrote the manuscript. P.L. and S.Z. assisted in the magnetoresistance measurement and data analysis. H.S. and H.Y. assisted in the data collection. Y.Z. supervised the magnetoresistance measurement and contributed to the data analysis and manuscript writing.

Competing Financial Interests Declared none.

doi:10.1038/nature12589

Zhang *et al.* reply

REPLYING TO J. Luo *et al.* *Nature* **501**, <http://dx.doi.org/10.1038/nature12589> (2013)

We agree with Luo *et al.*¹ that the magnetoresistance effects that we reported² were dependent on the method used to measure them. The reason that there is a difference in the results depending on whether method 1 or method 2 is used (adopting the measurement notation of

ref. 1) is that there are two voltage-stabilizing diodes in the Keithley 2400 instrument we used. We were unaware that when this instrument was used both as current source and voltmeter, one diode connected the input port of the current source to the input port of the

voltmeter, whereas the other diode connected the output port of the current source to the output port of the voltmeter. The diodes caused a crossover of the Hall coefficient from negative to positive when the instrument was used to conduct a Hall measurement in this configuration, leading us to propose an invalid mechanism for the abnormal magnetoresistance. Therefore the mechanism we proposed²—minority injection and an induced p-n boundary—does not provide a correct explanation for the observed geometry-enhanced magnetoresistance. Although such a mechanism does not operate in our samples, we note that a p-n boundary could still enhance magnetoresistance in certain circumstances according to our and others' theoretical calculations and experiments^{2–4}.

Once we became aware of the existence and role of the diodes, we were able to show that all the magnetoresistance properties reported previously² could be qualitatively reproduced by integrating two diodes between the indium electrodes within the devices themselves, resulting in magnetoresistance devices—consisting of the silicon wafer, the indium electrodes and now also the diodes—whose apparent properties are no longer dependent on the measurement methodology.

We are now developing a new mechanistic framework to describe this phenomenon, which takes into account the influences of both device geometry and the newly incorporated diodes. The mechanism and its generalization to other materials, (accompanied by corrections to our earlier invalid explanation) will be described in a forthcoming paper.

Finally, it is worth noting that (1) this magnetoresistance device could still be integrated into silicon-based electronics and therefore aid the development of magnetoelectronics, (2) the new mechanism

potentially provides ways to optimize device performance, such as improving the low-field sensitivity and reducing the power required and (3) our research² shows that the low-field magnetoresistance of $(\mu B)^2$ restricted by the Onsager relation⁵ could be enhanced in the nonlinear transport region, suggesting a way to improve the low-field sensitivity of semiconductors with only moderate mobility.

We thank Yang Ji at the Chinese Academy of Sciences, who initially alerted us to the methodology-dependence of our original results.

X. Z. Zhang^{1,2}, C. H. Wan^{1,2}, X. L. Gao^{1,2}, J. M. Wang^{1,2} & X. Y. Tan^{1,2}

¹Laboratory of Advanced Materials, Department of Materials Science and Engineering, Tsinghua University, Beijing 100084, China.

²Beijing National Center for Electron Microscopy, Tsinghua University, Beijing 100084, China.

email: xzzhang@tsinghua.edu.cn

1. Luo, J. *et al.* Low-voltage magnetoresistance in silicon. *Nature* **501**, <http://dx.doi.org/10.1038/nature12589> (2013).
2. Wan, C. H., Zhang, X. Z., Gao, X. L., Wang, J. M. & Tan, X. Y. Geometrical enhancement of low-field magnetoresistance in silicon. *Nature* **477**, 304–307 (2011).
3. Hu, J. S. & Rosenbaum, T. F. Classical and quantum routes to linear magnetoresistance. *Nature Mater.* **7**, 697–700 (2008).
4. Parish, M. M. & Littlewood, P. B. Non-saturating magnetoresistance in heavily disordered semiconductors. *Nature* **426**, 162–165 (2003).
5. Baker, D. R. & Heremans, J. P. Linear geometrical magnetoresistance effect: Influence of geometry and material composition. *Phys. Rev. B* **59**, 13927–13942 (1999).

doi:10.1038/nature12590